CODING FOR COOPERATIVE COMMUNICATIONS

A Dissertation

by

MOMIN AYUB UPPAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Electrical Engineering

CODING FOR COOPERATIVE COMMUNICATIONS

A Dissertation

by

MOMIN AYUB UPPAL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Zixiang Xiong
Committee Members,	Krishna Narayanan
	Gwan Choi
	Amarnath Banerjee
Head of Department,	Costas Georghiades

August 2010

Major Subject: Electrical Engineering

ABSTRACT

Coding for Cooperative Communications. (August 2010)
Momin Ayub Uppal, B.S., GIK Institute of Engineering Sciences and Technology;
M.S., Texas A&M University
Chair of Advisory Committee: Dr. Zixiang Xiong

The area of cooperative communications has received tremendous research interest in recent years. This interest is not unwarranted, since cooperative communications promises the ever-so-sought after diversity and multiplexing gains typically associated with multiple-input multiple-output (MIMO) communications, without actually employing multiple antennas. In this dissertation, we consider several cooperative communication channels, and for each one of them, we develop information theoretic coding schemes and derive their corresponding performance limits. We next develop and design practical coding strategies which perform very close to the information theoretic limits.

The cooperative communication channels we consider are: (a) The Gaussian relay channel, (b) the quasi-static fading relay channel, (c) cooperative multiple-access channel (MAC), and (d) the cognitive radio channel (CRC). For the Gaussian relay channel, we propose a compress-forward (CF) coding strategy based on Wyner-Ziv coding, and derive the achievable rates specifically with BPSK modulation. The CF strategy is implemented with low-density parity-check (LDPC) and irregular repeat-accumulate codes and is found to operate within 0.34 dB of the theoretical limit. For the quasi-static fading relay channel, we assume that no channel state information (CSI) is available at the transmitters and propose a rateless coded protocol which uses rateless coded versions of the CF and the decode-forward (DF) strategy. We implement the protocol with carefully designed Raptor codes and show that the im-

plementation suffers a loss of less than 10% from the information theoretical limit. For the MAC, we assume quasi-static fading, and consider cooperation in the low-power regime with the assumption that no CSI is available at the transmitters. We develop cooperation methods based on multiplexed coding in conjunction with rateless codes and find the achievable rates and in particular the minimum energy per bit to achieve a certain outage probability. We then develop practical coding methods using Raptor codes, which performs within 1.1 dB of the performance limit. Finally, we consider a CRC and develop a practical multi-level dirty-paper coding strategy using LDPC codes for channel coding and trellis-coded quantization for source coding. The designed scheme is found to operate within 0.78 dB of the theoretical limit.

By developing practical coding strategies for several cooperative communication channels which exhibit performance close to the information theoretic limits, we show that cooperative communications not only provide great benefits in theory, but can possibly promise the same benefits when put into practice. Thus, our work can be considered a useful and necessary step towards the commercial realization of cooperative communications.

To my Amma and Baba

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Zixiang Xiong, for his continuous support. I am extremely grateful to him for providing me with invaluable and inspirational guidance when needed, but at the same time providing me with a great degree of independence – he provided me with a perfect balance of the two. I would like to acknowledge the great support extended to me by our collaborators, Prof. Anders Høst-Madsen at University of Hawaii and Dr. Guosen Yue at NEC Laboratories America Inc. I am also thankful to some of the ex-members of the Multimedia Laboratory at Texas A&M University, Dr. Vladimir Stanković, Dr. Yong Sun, and Dr. Zhixin Liu, for their great help and valuable comments. I am grateful to Dr. Krishna Narayanan, Dr. Gwan Choi, and Dr. Amarnath Banerjee for serving on my dissertation committee.

Last, but not least, my heartiest appreciation to my wonderful wife, Juwairia, for her unconditional love and unfailing support.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Recent years have seen a growing trend in the use of wireless devices, with an ever increasing appetite for bandwidth intensive applications. In order to cope with the intricate demands put forth by these applications on current wireless networks, the concept of multiple-input multiple-output (MIMO) communications involving multiple tansmit/receive antennas has become extremely popular. Whereas MIMO communications promise great performance benefits, their drawback is the significant cost overhead associated with deploying multiple transmit/receive antennas. An alternative which holds immense economic promise is the idea of cooperative communications [1], which guarantees the ever so sought after spatial diversity gains typically associated with MIMO communications. Thus this technique is able to mimic the performance gains of MIMO systems *without* actually employing multiple antennas. As a result, the area of cooperative communications has generated tremendous research interest in recent years. Some works in this area include [1, 2, 3, 4, 5] amongst a host of others. In the following, we will first give a brief layman's introduction to the idea of cooperative communications. We will then describe the summary of the dissertation and explain how it has been organized.

A. Introduction to Cooperative Communications

Loosely speaking, cooperative communications is based on grouping closely located networks nodes in wireless networks together into clusters, inside which the nodes cooperate when sending and/or receiving information – in this way, different clusters

---

´The journal model is *IEEE Transactions on Automatic Control.*

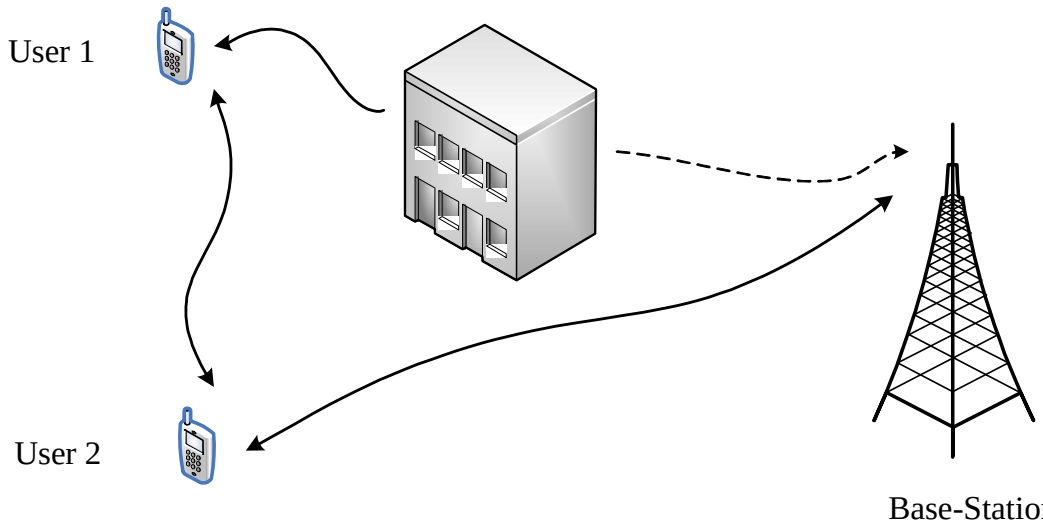


Fig. 1. Cooperative communication channel with two users and a single base-station. The channel from user-1 to the base-station is relatively weak.

act as large transmit and/or receive antenna arrays. A simple form of a cooperative communication network is shown in Fig. 1 where two users wish to communicate their individual messages to a single base-station. The figure shows a specific scenario where the channel from the first user to the base-station is very noisy – noisy to an extent where it cannot support the required data rate. On the other hand, the channel from the second user to the base-station is quite strong. If the second user channel is strong enough, User-1 can seek User-2's cooperation in relaying its message to the base-station. Thus, when User-1 transmits to the base-station, User-2 overhears the transmission and then forwards User-1's message to the base-station along with its own. It is quite evident that this model is a cooperative version of a multiple-access channel (MAC). The channel model in Fig. 1 can also be viewed as a cooperative broadcast channel where the base-station intends to send different messages (or the

same message in case of a cooperative multi-cast channel) to the two users and the users cooperate amongst each other when receiving the information.

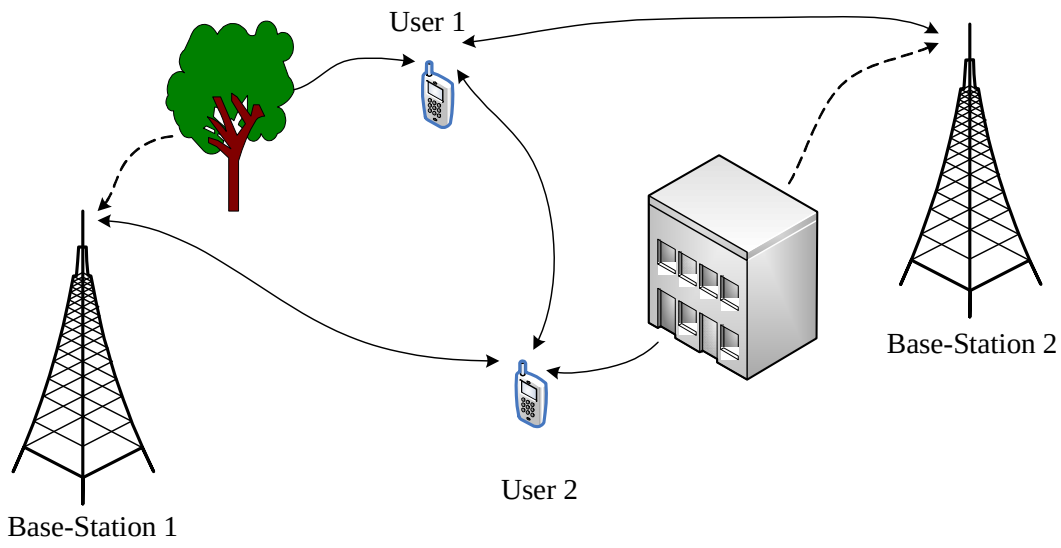


Fig. 2. Cooperative communication channel with two users that wish to communicate with their respective base-stations. A user sees a relatively weaker channels to its own base-station compared to that to the other user's base-station.

An extension of the cooperative communication channel of Fig. 1 is shown in Fig. 2 where the two users intend to transmit their distinct messages to their respective base-stations. However, for the specific condition shown in the figure, each user sees a relatively weak channel to its own base-station. As a result, the users are unable to communicate directly with the base-stations at the required data rates. However, note that a user sees a sufficiently strong channel to not only the other user, but also the other user's base-station. Thus the cooperative communication strategy would dictate that the users exchange their messages first and each user relay the other user's message to its respective base-station. Such a cooperation channel is termed as the transmitter cooperative interference channel [3]. Note that a similar cooperation

strategy also applies when the base-station intends to transmit some messages to their respective users where the users cooperate amongst themselves when receiving information. Such a cooperation channel is referred to as the receiver cooperative interference channel [3].

B. Summary of Contributions

As mentioned before, cooperative communications promises, at least in theory, significant performance gains over a traditional non-cooperative system. However, cooperation cannot be considered a viable option unless it can, when put in practice, obtain gains comparable to those promised by theory. In this dissertation, we intend to explore the feasibility of cooperative communications by developing practical coding strategies and comparing simulated practical performance to the *derived* theoretical limits. Thus the objectives of this dissertation are two-fold:

a) to develop information-theoretic coding schemes for several cooperative communication channels and derive the corresponding achievable rates/performance limits, and

b) to design and implement practical coding schemes which follow the spirit of the information-theoretic analysis and to evaluate whether such practical cooperation schemes still yield the gains over non-cooperative strategies.

Keeping in mind the objectives mentioned above, we will study four types of cooperative communication channels which are listed below.

1. The Gaussian relay channel

2. The quasi-static fading relay channel

3. The cooperative multiple-access channel

4. The cognitive radio channel

In the following, we give brief overviews of our contributions in regards to each one of these channels. These introductory remarks have been kept to a minimum. More introductory information about each one of these channels can be found in the subsequent chapters.

1. The Gaussian relay channel

Perhaps the simplest form a cooperative communication network is a three-node relay channel [6]. One can view the communication model in Fig. 1 as a relay channel if one of the users does not have any information to transmit, and the sole purpose of which is to help the other user in its transmission. Even with this simplest setup, relaying promises significant cooperation gains over traditional point-to-point communication. In this dissertation, we will first study compress-forward (CF) coding with binary phase-shift keying (BPSK) for the half-duplex Gaussian relay channel [7, 8]. In CF relaying, Wyner-Ziv coding is applied at the relay to exploit the joint statistics between signals at the relay and the destination. We propose Slepian-Wolf coded nested scalar quantization (SWCNSQ) for practical Wyner-Ziv coding at the relay. After providing the achievable rate of SWCNSQ based CF relaying as a performance benchmark, we will present a practical code design using low-density parity-check (LDPC) codes for error protection at the source, and nested scalar quantization plus irregular repeat-accumulate (IRA) codes for CF coding at the relay. The degree distributions of the LDPC and IRA codes are optimized using extrinsic information transfer charts and a Gaussian approximation. Under discretized density evolution for asymptotically large block lengths, our optimized code design operates $0.11 - 0.21$ dB

away from the SWCNSQ limit for CF relaying. Simulations with LDPC/IRA codes of length 2×10^5 bits show a performance gap of $0.27 - 0.38$ dB from the achievable rate.

2. The quasi-static fading relay channel

As opposed to the Gaussian relay channel, we also consider the case of a half-duplex wireless relay channel where all links experience independent quasi-static Rayleigh fading and where the instantaneous channel realizations are unavailable at the transmitters [9, 10]. We assume that the network does not have a stringent delay constraint – thus the source and/or the relay continue transmitting until the destination acknowledges successful decoding. We identify rateless coded relaying as the natural choice, and propose a rateless coded protocol where each transmission from the source and/or the relay adds incremental redundancy to help the destination recover the original message. Our proposed protocol utilizes, in conjunction with rateless coding, a combination of the two popular relay cooperation schemes, namely decode-forward and CF. Assuming very limited feedback from the destination, we derive the theoretical achievable rates specifically with BPSK. We then implement the rateless coded relaying protocol using Raptor codes. The degree profiles for the Raptor codes are designed to maximize the average throughput − with the design formulated as a convex optimization problem. Using discretized density evolution for asymptotically large block lengths, the optimized codes lose approximately 5% in performance from the theoretical limit, whereas with practical finite block lengths, the performance loss is approximately 9%.

3. The cooperative multiple-access channel

A natural extension to the relay channel is the case of a cooperative multiple-access channel (MAC), shown in Fig. 1, where the role of relaying is played by other users who also have their own information to transmit. In this dissertation, we will consider cooperation in the low power (low SNR) regime of the MAC with the assumption that the transmitters have no channel state information [11, 12, 13]. A relevant performance to consider is therefore the outage capacity. We develop cooperation methods based on *multiplexed coding* in conjunction with rateless codes and find the achievable rates and in particular the minimum energy per bit required to achieve a certain outage probability. We consider two modes of cooperation: full duplex (code division multiple access or CDMA), where nodes can transmit and receive simultaneously on the same frequency band, and half duplex (frequency division multiple access or FDMA), where the nodes transmit and listen on different frequency bands. We show that, perhaps surprisingly, there is little loss in performance when using FDMA. Furthermore, our results show that multiplexed rateless codes come within 0.1 dB of the outer bound on capacity. We also develop practical rateless coding methods for FDMA using multiplexed Raptor codes which operate within 0.52 and 1.1 dB of the theoretical limit for the two- and four-user case, respectively.

4. The cognitive radio channel

We implement a dirty-paper coded framework for the cognitive radio channel (CRC) [14]. We assume that the cognitive user has non-causal knowledge about the primary user's transmissions. Thus the secondary receiver can employ dirty-paper coding to counter the effect of any interference from the primary user. In addition, we consider a situation where the introduction of the cognitive user should not affect

the performance of the primary system – nor should the primary system have to change its encoding/decoding process. For the primary user we use an LDPC code and a 4-ary pulse amplitude modulation format. For the cognitive user, we propose a multi-level dirty-paper coding scheme which employs trellis-coded quantization for source coding and LDPC codes for channel coding. At a transmission rate of 1.0 bits/sample, the designed dirty-paper coding scheme operates within 0.78 dB of the theoretical limit, which we believe is the best performance reported in the literature for this rate.

C. Notation

Notation-wise, all logarithms are of base two unless otherwise stated; vectors and matrices are represented by boldface letters, with their dimensions indicated by the context in which they are used. $|\cdot|$ denotes magnitude of a complex number and $||\cdot||$ represents norm of a vector.

D. Dissertation Organization

The dissertation is organized as follows. In Chapter II, we discuss some introductory concepts related to source coding with side-information as well as channel coding with side-information, which will be used in the code designs in the chapters to follow. The next four chapters form the main body of the dissertation, with each chapter devoted to one of the four cooperative communication channels mentioned above. Specifically, we discuss the Gaussian relay channel in Chapter III, the quasi-static relay channel in Chapter IV, the cooperative MAC in Chapter V, and the CRC in Chapter VI. Finally, we provide the concluding remarks in Chapter VII.

CHAPTER II

CODING WITH SIDE-INFORMATION

In this chapter, we provide brief introductions to the concept of coding with side-information, which plays a major role in the code designs for several cooperative communication channels, as will be shown in the ensuing chapters. In general, coding with side-information can be divided into two categories. The first category is termed source coding with side-information (SCSI), where one needs to compress a source while accounting for correlated side-information available at the decoder but not at the encoder. The second category is the dual of SCSI, called channel coding with side-information (CCSI), where the encoding accounts for the presence of interference which is available as side-information at the encoder but not at the decoder. In the following, we will briefly discuss basic concepts related to these two categories.

A. Source Coding with Side-information

In Chapters III and IV, we will use the concept of SCSI to develop coding strategies for the Gaussian relay channel, and the quasi-static fading relay channel, respectively. The basic system setup for SCSI is shown in Fig. 3. The source wishes to compress the sequence $\mathbf{X}$ which is to be communicated to the destination over a noiseless channel. However, the destination has available with it the sequence $\mathbf{Y}$ which is correlated with the information sequence $\mathbf{X}$. The source-coding or compression can be lossless, where the recovered sequence $\hat{\mathbf{X}}$ has to be equal to the information sequence with arbitrarily small probability of error. It can also be lossy, where the recovered sequence $\hat{\mathbf{X}}$ is allowed to be a distorted version of $\mathbf{X}$, while satisfying a certain fidelity criteria. When the sequence is drawn from a discrete alphabet, the lossless compression under the SCSI setting is referred to as Slepian-Wolf (SW) coding [15]. On the other hand,

lossy compression in the SCSI setup is typically referred to as Wyner-Ziv (WZ) coding [16]. In the following, we discuss the two separately by first providing the information theoretic bounds, followed by some short discussions on ways of implementation.

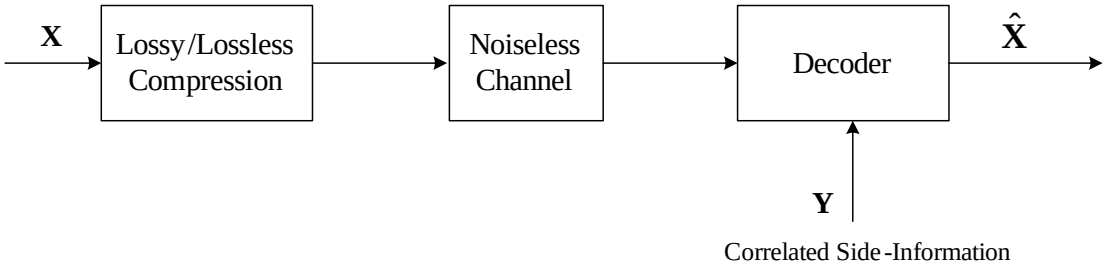


Fig. 3. Source coding with side-information. The side-information sequence is correlated with the information sequence and is available at the decoder but not at the encoder.

1. Slepian-Wolf coding

Consider two infinite length sequences $\mathbf{X}$, $\mathbf{Y}$ which are drawn i.i.d. from a pair of correlated, discreet random variables X and Y. If the two sequences are to be jointly compressed, a compression rate of $R_X + R_Y = H(X,Y)$ is sufficient to guarantee that the recovered sequences when decoded jointly are the same as the original sequences with arbitrarily low probability of error [17]. What if the two sequences are encoded separately, with the encoder for one sequence having no knowledge of the other sequence? Slepian and Wolf [15] proved the surprising result that with separate encoding but joint decoding, a rate of $R_X + R_Y = H(X,Y)$ is still sufficient to correctly recover the two sequences. The model shown in Fig. 3 is a specific case of this setup (corresponding to a corner point on the Slepian-Wolf rate region), where $\mathbf{Y}$ is already known at the decoder. Under the setting of Fig. 3, the Slepian-Wolf result translates into the fact that the rate required for lossless compression of $\mathbf{Y}$ is $R_Y = H(Y)$, and therefore the required rate for lossless compression of $\mathbf{X}$ with $\mathbf{Y}$

available as the decoder side-information is given as

$$R_X = H(X,Y) - R_Y = H(X|Y).$$

This result is quite surprising since it states that the non-availability of $\mathbf{Y}$ at the encoder does not affect the compression rate – one can guarantee lossless compression using the same rate as if the side-information was also available at the encoder.

a. Practical Slepian-Wolf coding

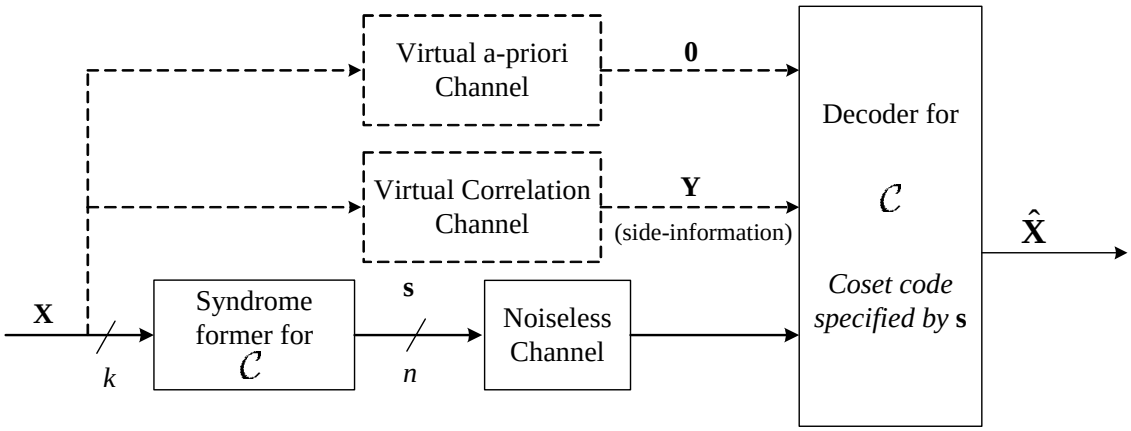


Fig. 4. Slepian-Wolf coding with parity-check (channel) codes.

The information theoretic proof of the Slepian-Wolf theorem [15] is based on random binning arguments, which is not possible to put into practice. In reality, one has to follow a structured binning approach with algebraic operations [18]. An approach to structured binning is through the use of parity-check codes as first suggested by Wyner in his 1974 paper [19]. The basic idea behind the use of parity-check codes to approach the corner points in the Slepian-Wolf rate region (when sequence $\mathbf{Y}$ is known as side-information at the decoder) is shown in Fig. 4. Compression of a length k sequence $\mathbf{X}$ involves mapping the sequence to its corresponding n-bit syndrome. Thus the compression rate is given as $R_{SW} = \frac{n}{k}$ and is related to the rate

R of the parity-check code as

$$R = \frac{k-n}{k} = 1 - R_{SW}.$$

The syndrome sequence, which identifies the coset of the parity-check code to be used for decoding, is transmitted to the destination over the noiseless channel. In addition, since the side-information $\mathbf{Y}$ is correlated with the information sequence $\mathbf{X}$, it can be thought of as being transmitted over a virtual correlation channel with $\mathbf{X}$ as the input, and $\mathbf{Y}$ as the output. The capacity over this virtual correlation channel is given as $C_{corr} = I(X;Y)$. In addition, if the sequence $\mathbf{X}$ is not equally likely, one can think of $\mathbf{X}$ being transmitted over a virtual a-priori channel which provides information $C_{ap} = 1 - H(X)$, where all logarithms are assumed to be to the base equal to the alphabet size of $\mathbf{X}$ and $\mathbf{Y}$ – hence the term 1 in $C_{ap} = 1 - H(X)$. Using an informal argument, one can show that if the parity-check code is capacity-achieving over the joint virtual correlation and a-priori channels, one can achieve the Slepian-Wolf limit using the setup in Fig. 4. The argument goes as follows. If the parity-check code is capacity achieving, then $\mathbf{X}$ can be recovered error free if the rate R of the code is less than the overall capacity on the two virtual channels, i.e.,

$$\begin{aligned} R &\leq C_{corr} + C_{ap} = I(X;Y) + 1 - H(X) \\ &= 1 - H(X|Y) \end{aligned} \tag{2.1}$$

Now substituting $R = 1 - R_{SW}$ in (2.1), we obtain the condition for error free recovery of $\mathbf{X}$ as

$$R_{SW} \geq H(X|Y), \tag{2.2}$$

which is the same as the SW compression limit.

Over the past few years, several works have appeared in the literature related to practical designs for Slepian-Wolf compression. Xiong *et. al* [20, 21] were probably the first to follow the binning scheme mentioned above in designing SW coding schemes based on turbo/low-density parity-check (LDPC) codes – the design with two binary symmetric sources performs only 0.0389 bits away from the SW coding limit. Some other related works include, but are not limited to [22, 23] and [24].

2. Wyner-Ziv coding

As mentioned before, WZ coding refers to the lossy compression of $\mathbf{X}$ in Fig. 3 with a certain fidelity criteria, under the condition that the side-information $\mathbf{Y}$ is available at the decoder but not at the encoder. The problem, also referred to as rate-distortion with side-information was investigated in [16]. The objective is to determine how many bits $R_{WZ}(D)$ are necessary to encode $\mathbf{X}$ such that the decoder is able to obtain a reconstruction $\hat{\mathbf{X}}$ which has a distortion of at most D from the original sequence, i.e. $\mathbb{E}\left[d(X,\hat{X})\right] \leq D$, where $d(\cdot,\cdot)$ is an arbitrary distortion metric. According to [16], the rate-distortion function $R_{WZ}(D)$ is given as

$$R_{WZ}(D) = \inf_{f(W|X)} \inf_{F} \left[I(X;W) - I(Y;W)\right], \tag{2.3}$$

where W is an auxiliary random variable such that $Y \rightarrow X \rightarrow W$ forms a Markov chain. As a result of the Markov chain, we have $I(X;W) - I(Y;W) = I(X;W|Y)$. The infimum in (2.3) is taken over all random variables W and over all reconstruction functions $\hat{X} = F(Y,W)$, which satisfy the distortion constraint given by

$$\mathbb{E}\left[\, d\left(X, F(Y,W)\right) \,\right] \leq D, \tag{2.4}$$

where the expectation is with respect to the joint pdf of X, Y, and W. It was also shown in [16] that under the setup described above, one can do not better than the

limit in (2.3).

In general, the rate-distortion function with side-information in (2.3) suffers a rate loss compared to the case where the encoder also has information about the side-information $\mathbf{Y}$. However, when X and Y are jointly Gaussian with zero mean, and with mean square error (MSE) as the distortion metric, the rate-distortion function $R_{WZ}(D)$ is the same as that for the case where $\mathbf{Y}$ is also available at the encoder. Indeed, if one chooses $W = X + Z$, where Z is a zero-mean Gaussian random variable whose variance is chosen to satisfy the MSE distortion constraint, and with the reconstruction function $F(\cdot, \cdot)$ equal to the conditional expectation of X given W and Y, one can show that [25]

$$R_{WZ}(D) = \frac{1}{2} \log^+ \frac{(1-\rho^2)\sigma_x^2}{D}, \tag{2.5}$$

where ρ is the correlation coefficient between X and Y, σ_x^2 is the variance of X and $\log^+(x) = \log(\max(x, 1))$. It can be verified that this rate-distortion function is the same as that with joint encoding at the source [25]. In practice, many image and video sources can be modeled as jointly Gaussian. Therefore, the result stated above is important since it promises no loss in coding efficiency when the side-information is available at the decoder only, and not at the encoder.

a. Practical Wyner-Ziv coding

The Gaussian example presented in the previous subsection gives us an insight into practical WZ coding. Basically, the auxiliary random variable W can be thought of as the output of a quantizer with X as the input. At the decoder, one needs to estimate the sequence $\mathbf{X}$ given the quantized output $\mathbf{W}$ and the side-information sequence $\mathbf{Y}$. From an information theoretic perspective, there are granular gain and boundary gain in source coding, while shaping gain and packing gain in channel coding [26].

Since WZ coding is originally a source coding problem, one needs to consider the granular and boundary gains. However, as opposed to traditional source coding, the side-information necessitates compression with binning, which can be accomplished by channel coding, as pointed out in Section A. This channel coding however is not conventional in the sense that there is only packing gain, but no shaping gain. One can easily draw equivalence between the boundary gain in source coding and the packing gain in channel coding. Hence in WZ coding, the granular gain is achieved through source coding or quantization, and the boundary gain via syndrome based compression through channel coding. In short, even though WZ coding is inherently a source coding problem, its practical implementation involves both source and channel coding components.

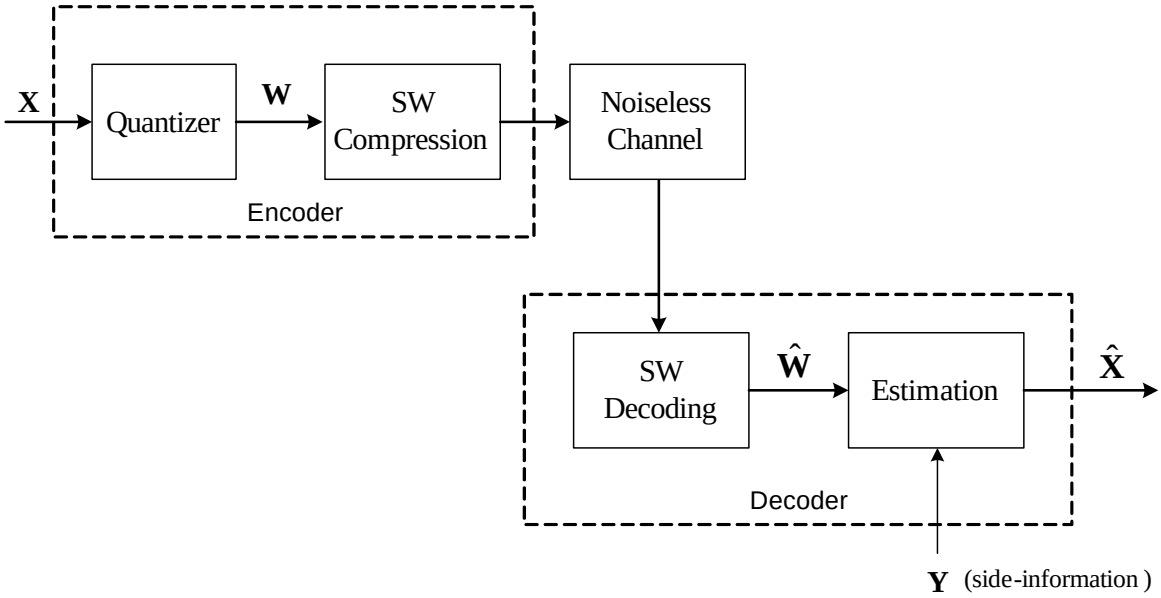


Fig. 5. Practical WZ coding with quantization and SW coding.

A source-channel coding strategy for WZ coding using nested lattices was proposed in [18] and was shown to achieve the WZ limit asymptotically as the dimensionality of the lattice codes approaches infinity. However, the implementation of

such high dimensional lattice codes is not practical. A somewhat related, but practical approach to WZ coding is shown in Fig. 5, where the input sequence is first quantized to obtain the WZ auxiliary random variable. In general, the quantized output is discrete which is still correlated with $\mathbf{Y}$. Thus, in practice, one can exploit this correlation to compress the quantized output further using SW coding, which is implemented through channel codes as mentioned in Section A. In order to reach the theoretical WZ coding limit in practice, one needs to employ a good source code, e.g. trellis coded quantization (TCQ) [27], which can achieve as high a granular gain as possible. At the same time, one needs to employ capacity approaching channel codes, e.g. turbo and LDPC codes that can achieve the Slepian-Wolf limit.

Finally, one can easily draw parallels between classical entropy constrained source coding (quantization) and the practical WZ coding approach in Fig. 5. The only difference between the two is that the latter needs to account for the side-information which is available at the decoder. No wonder, if one replaces the SW coding (conditional entropy coding) in Fig. 5 with classical entropy coding, one obtains the classical entropy constrained quantization problem. For this reason, the scheme in Fig. 5 can be referred to as Slepian-Wolf coded quantization (SWCQ). For the quadratic Gaussian Wyner-Ziv problem, where X and Y are jointly Gaussian, it was shown in [28] that at high rates, the performance gap of SWCQ scheme to the WZ distortion-rate function $D_{WZ}(R)$ is exactly the same as the performance gap between classical entropy constrained quantization and the classical distortion-rate function $D(R)$. In a practical design example [29] of SWCQ with 2-D TCQ for quantization, irregular LDPC codes for SW coding, MSE as the distortion measure, and optimal conditional mean estimation at the decoder, the performance gap to WZ distortion-rate function was reported to be only 0.66 dB at 1.0 bits/sample (b/s) and 0.47 dB at 3.3 b/s.

B. Channel Coding with Side-information

CCSI, as mentioned before, refers to the problem of communicating over a noisy channel with some knowledge of the channel state available as side information at the encoder, but not at the decoder. Gelfand and Pinsker [30] obtained the capacity for the problem involving a discrete memoryless channel in 1980. Three years later Costa [31] used Gelfand's and Pinsker's result to formulate the theory for the special case of Gaussian channel. Costa's work, also referred to as "writing on dirty paper", did not address the relevance of its results to communication networks and hence did not draw much attention at first. However, we now know that several situations in communication networks can be modelled as a CCSI problem e.g. ISI channels, cross talk interference pre-subtraction in vectored digital subscriber line, broadcast channels, cognitive user channels, and transmitter cooperative networks to name a few. Moreover, CCSI also finds widely celebrated applications in covert operations such as data hiding and watermarking.

In Chapter VI, we will discuss code designs for a Gaussian cognitive radio channel, an enabling component of which is CCSI. Since dirty-paper coding (DPC) is a specific to Gaussian channels, we will mostly discuss DPC as a special case of CCSI in this section. The objective of this section is to first review some theoretical aspects of DPC. This will be followed by discussion of some guidelines for developing practical DPC strategies, which we will use when we develop a practical DPC scheme for the cognitive radio channel. The organization of this section is as follows. In Section 1 we will introduce the Gelfand-Pinsker coding problem and discuss how it applies to the special case of Costa coding. Section 2 discusses guidelines to developing practical approaches for solving the DPC problem, which will highlight the importance of source coding in the apparent channel coding problem of DPC. Finally, Section 3

will present an information theoretic perspective to the requirement of a source code in DPC.

1. Gelfand-Pinkser coding and Costa coding

Gelfand and Pinsker [30] considered the case of CCSI in a discrete memorlyless channel. The channel model is shown in Fig. 6. The input to the channel is denoted by

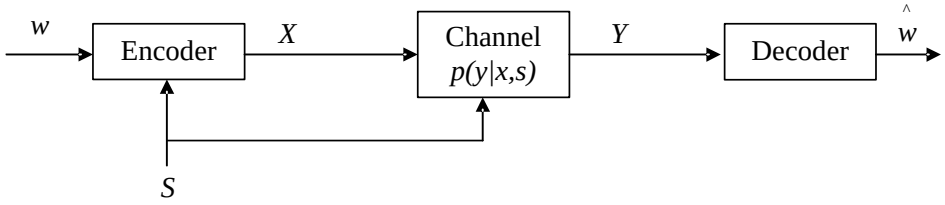


Fig. 6. Gelfand-Pinsker Channel

X, the output by Y, and the side information by S which is known non-causally at the encoder but not at the decoder. The encoder is to transmit message w over a discrete memoryless channel characterized by the transition probability $p(y|x,s)$. It was shown in [30] that the capacity of this channel is given by

$$C = \max_{p(v,x|s)} \big(I(V;Y) - I(V;S)\big), \tag{2.6}$$

where V is an auxiliary random variable. The proof of Gelfand-Pinsker capacity is based on random coding and binning. For the general CCSI, Gelfand-Pinsker coding suffers a loss compared to the situation when the side information is available at both the encoder and the decoder.

Costa [31] used the general formula in (2.6) to prove the capacity of a Gaussian channel, where the signal is corrupted by an additive Gaussian noise as well as Gaussian interference – Costa's channel model is shown in Fig. 7. Costa drew an analogy of this channel to the problem of writing on a sheet of paper covered with dirt, where

the writer knows the location and intensity of the dirt particles but the reader does not. Thus the whimsical title of "dirty-paper coding".

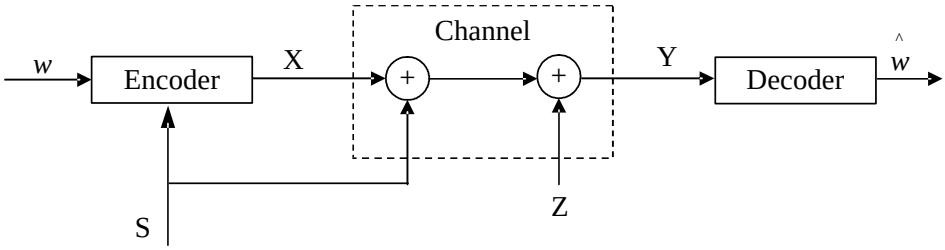


Fig. 7. Costa channel

The transmitter wishes to send the message such that a power constraint $\mathbb{E}[|X|^2] \leq P_t$ is satisfied. The output of the channel is given by $Y = X + S + Z$, where the interference $S \sim \mathcal{N}(0, P_Q)$ is known non-causally at the transmitter but not at the receiver, and $Z \sim \mathcal{N}(0, P_Z)$ is the additive noise. If the auxiliary random variable is chosen as $V = X + \alpha S$, with $\alpha = \frac{P_t}{P_t+N}$, Costa proved the surprising result that the capacity of the channel in Fig. 7 is the same as if the interfering signal S were not present at all. In other words, the Gelfand-Pinsker capacity in (2.6) is the same as that of the typical AWGN channel, and is given as

$$C = \frac{1}{2} \log \left(1 + \frac{P_t}{P_Z} \right). \tag{2.7}$$

Costa's proof is once again based on random coding and binning arguments. Although Costa proved this result for a Gaussian interference, it was later generalized for any arbitrary distribution on S in [32].

2. Approaches to practical DPC

Since Costa's proof is based on random coding and binning, its practical implementation is not possible. However, it does provide a very visible clue of "binning". Not surprisingly, many recent works on practical schemes for DPC e.g. [33, 34, 35] have utilized the concept of structured binning and obtained performance close to the Costa's limit. For example, [33] focused on the high rate regime and designed a DPC scheme based on nested turbo codes which was able to perform within 1.42 dB of the capacity at a rate of 1.0 b/s. On the other hand, [34, 35] developed DPC schemes based on TCQ and LDPC/IRA codes for the low rate regime. At a transmission rate of 0.25 b/s, the scheme in [35] performs 1.3 dB away from capacity. On the other hand, the authors in [34] were able to devise a scheme which performs only 0.63 dB away from the theoretical limit at a transmission rate of 0.25 b/s.

In this section we will introduce Tomlinson-Harashima precoding (THP) which can be seen as a one dimensional implementation of DPC. We will then draw parallels between THP and scalar quantizers, and thus show the need of a source code in solving the DPC problem. Finally, we will introduce a structured binning strategy based on nested lattices [18].

a. Tomlinson-Harashima precoding

THP [36, 37] shown in Fig. 8 was originally designed to counter the interference in ISI channels. Consider a message codeword U to be transmitter over a channel characterized by an additive interference S and an additive noise Z, with powers P_Q and P_Z, respectively. The interference S is available non-causally to the encoder but not to the decoder. One can immediately see the equivalence of this problem to DPC if the noise Z were Gaussian. At first glance one would consider pre-subtracting the

side information from the transmitted signal in order to cancel the interference, i.e., transmitting $X^s = U - S$. Indeed, the received signal will now be $Y^s = X^s + S + Z = U + Z$, and hence interference free. A closer look at this approach however reveals that this pre-subtraction would have to pay a severe power penalty. Assuming that U and S are independent, the transmitter power will be $\mathbb{E}[|X^s|^2] = \mathbb{E}[|U|^2] + \mathbb{E}[|S|^2]$. Since the side information can have an arbitrarily high power, $\mathbb{E}[|X^s|^2]$ can be much higher than $\mathbb{E}[|U|^2]$, which will result in a severely reduced transmission rate than (2.7). In order to avoid this power penalty, THP uses *modulo* arithmetic in order to constrain the transmitted signal to a finite interval.

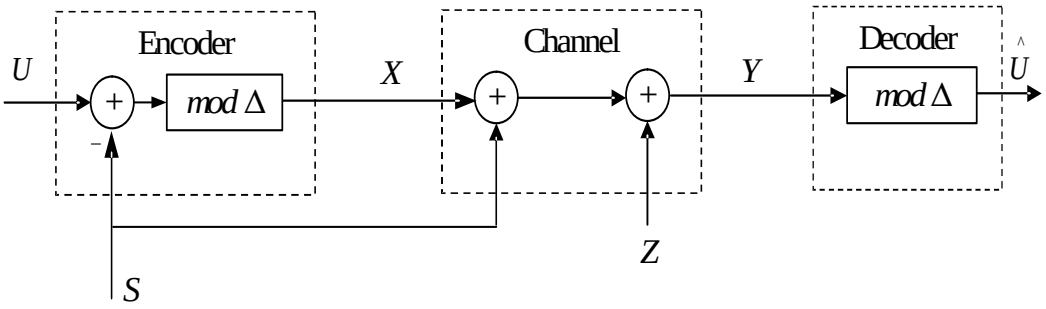


Fig. 8. Tomlinson-Harashima precoding

Let the codeword to be transmitted U be constrained to a finite interval of length Δ, i.e., $U \in [0, \Delta]$. The signal transmitted to the channel is $X = (U - S)\ mod\ \Delta$. Because of the *mod* operation, X is now limited to the same finite interval as U and hence it does not suffer the power penalty which a simple pre-subtraction would. At the decoder, a same *mod* operation is performed to get an estimate of U. In the absence of noise, THP guarantees that U is recovered without error at the decoder.

This can be shown as follows. The recovered codeword $\widehat{U}$ is given by

$$\begin{aligned}
\widehat{U} &= Y \ mod \ \Delta \\
&= \big(X + S\big) \ mod \ \Delta \\
&= \big((U - S) \ mod \ \Delta + S\big) \ mod \ \Delta \\
&= \big(U - S + S\big) \ mod \ \Delta \\
&= U \ mod \ \Delta \\
&\overset{(a)}{=} U,
\end{aligned}$$

where (a) follows from the fact that $U \ \in \ [0, \Delta]$.

b. THP with scalar quantizers

The encoding process in THP reduces the signal $U' = U - S$ to one of the equivalent representatives of the symbol given as $n\Delta$, where $n = \lfloor \frac{U'}{\Delta} \rfloor$. The difference $X = U' - n\Delta$ is then transmitted to the channel. One can draw parallels between the output of the *mod* operation in THP and the quantization error in a scalar quantizer. Consider a scalar uniform quanitzer whose quantization points are given by $n\Delta$ with $n \in \mathbb{Z}$. If U is distributed on the interval $[0, \Delta)$, then the *mod* operation in THP is related to the quantizer by

$$U' \ mod \ \Delta = U' - Q\big(U' - \frac{\Delta}{2}\big),$$

where $Q\big(\ \cdot\ \big)$ represents uniform quantization over a constellation with points $n\Delta$. It can be shown that the *mod* operations in THP can be replaced by the scalar quantizer by making sure that the input signal is distributed on the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ instead of on $[0, \Delta)$. Fig. 9 shows equivalent THP with scalar quantizers. When the interference power P_Q is large, the quantization error X is approximately uniformly distributed on the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ and hence the power of the transmitted signal is independent

of P_Q and is approximately given by $\frac{\Delta^2}{12}$.

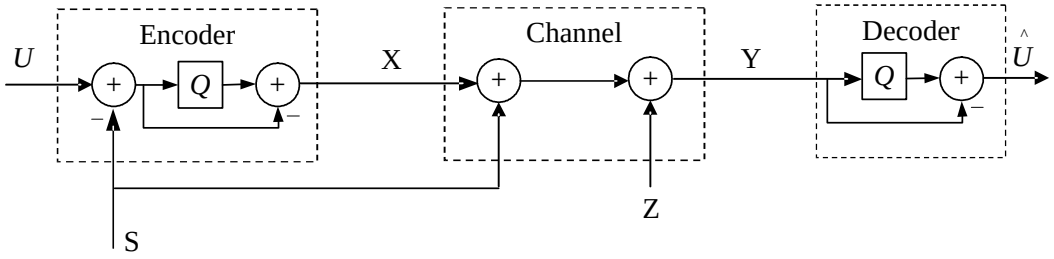


Fig. 9. Tomlinson-Harashima precoding with scalar quantizers

c. Generalization of THP to vector quantizers

As pointed out in [32, 38] THP suffers a signigicant loss from Shannon's capacity limit, especially at low signal to noise ratios (SNRs). The main drawback of THP is that it only uses the current value of the side information S and does not consider the future values. The *mod* operation is performed on a symbol by symbol basis resulting in an output which is uniformly distributed on $[-\frac{\Delta}{2}, \frac{\Delta}{2})$. This is equivalent to performing a *mod* operation over a high dimensional cuboid, which suffers a *shaping loss*. An optimal quantizer however should be equivalent to performing a *mod* operation over a high dimensional sphere, resulting in Gaussian quantization error in each dimension. Thus instead of using the side information on a symbol by symbol basis, one needs to consider an entire sequence. The solution to recovering the shaping loss therefore lies in performing a high dimensional *mod* operation, or equivalently in *vector quantization*.

d. Binning based on nested lattices

So far, we have only discussed the source coding (quantization) portion of the DPC problem, which is essential to satisfy the power constraint. We found that one can accurately retrieve the coded message in the absence of noise. However, in practice one needs to add error protection to the transmission in order to combat the channel's additive Gaussian noise. This therefore introduces an additional channel coding aspect to the problem. The question here is: How do we view the joint source and channel code design under a similar framework? Zamir *et al* [18] proposed a practical binning scheme based on nested codes. Hence the solution to the Gelfand-Pinsker problem lies in nested parity check codes, and in nested lattice codes for the Costa coding problem.

A nested lattice code comprises of a *coarse lattice code* nested inside a *fine channel code*, i.e., every codeword of the coarse lattice code is also a codeword of the fine lattice code but every codeword of the fine lattice is not a codeword of the coarse lattice. According to [18], for a good dirty-paper code design, the fine code should be a good channel code whereas the coarse code should be a good source code. Hence the source code is nested within the channel code. The concept of binning can be derived from this nesting approach. The group of channel codewords nested within a single source codeword are said to belong to the same *bin*, where the bin is indexed by that particular source codeword.

Let us illustrate how binning based on nesting works by considering a one dimensional nested lattice as an example. Note that we select a one dimensional lattice for illustrative purposes only – in practice a high dimensional lattice should be used in order to achieve good performance. Fig. 10 demonstrates a binning strategy based on a 1-D nested integer lattice. The points on the lattice indexed by a 0 correspond to

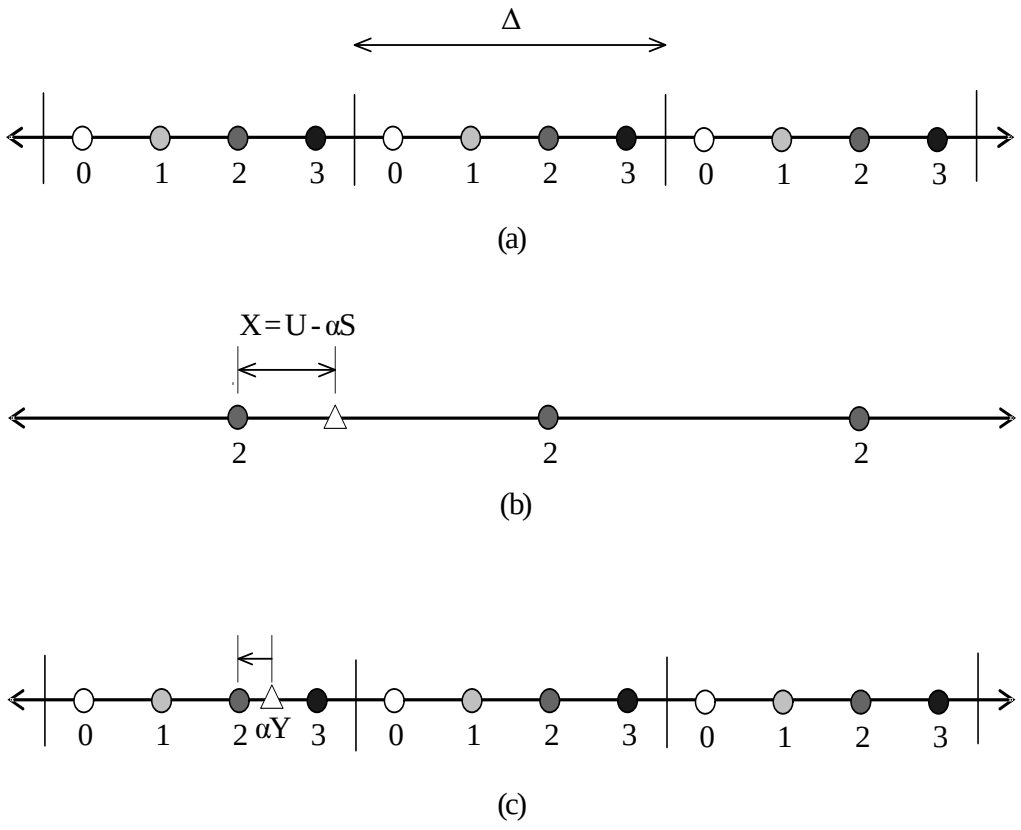


Fig. 10. Binning scheme using a 1-D nested lattice (a) Nested lattice (b) Encoding (c) Decoding

the channel codewords in the basic coset. Similarly the points indexed by the other numbers correspond to the other cosets. The message to be transmitted (which in this case will be a two bit message) selects one of these cosets. In this example, coset 2 is indexed by the message. The message is first scaled by a factor α (the necessity of this scaling comes from Costa's original proof in [31]). This scaled side information is then quantized to the nearest codeword in the coset 2 and the quantization error is sent to the channel. At the decoder the nearest codeword to the scaled received signal is found to get an estimate of the transmitted signal. The decoded message therefore is the index of this estimate. THP with scalar quantizers can be viewed as a binning scheme based on nested lattices. The input U in Fig. 9 can be thought of

as a channel codeword selected by the message. Quantizing the difference $U - \alpha S$ to an infinite integer lattice $n\Delta$ is the same as quantizing αS to a lattice where the channel codeword U has been infinitely replicated. In practice, this nested code strategy can be implemented by employing the likes of TCQ as the source coding, and LDPC/Turbo codes as the channel coding component.

3. Information theoretic perspective of lattice strategies

An information theoretic framework for studying the Costa coding problem was presented in [32]. Costa coding is inherently a channel coding problem. According to [26], there are packing and shaping gain in channel coding. The shaping gain has to do with the shape of the Voronoi region of the lattice, which ideally has to be a sphere. The packing gain has to do with the way the code regions are packed against each other. Costa coding problem as explained earlier can be split into a source coding and channel coding component. The source coding becomes necessary to satisfy the power constraint and is hence required to reduce the scaled side information modulo the Voronoi region. The constellation therefore needs to be replicated infinitely so that one can quantize the side information to satisfy the power constraint. This source coding therefore is not conventional in the sense that it only has the granular gain and no boundary gain. One can easily draw equivalence between the granular gain in source coding and the shaping gain in channel coding. Hence in channel coding with side information problem, the shaping gain is achieved through source coding and the packing gain through channel coding. In order to get close to the Costa capacity limit, the source coder should be designed such that its Voronoi region is almost a spherical region in high dimensional Euclidean space (such as TCQ). Similarly the channel code should also be near capacity (such as Turbo codes or LDPC).

Erez et al [35] provided an information theoretic limit of DPC when employing

lattice precoding. Let Λ be an L dimensional lattice quantizer with the basic Voronoi region $\mathcal{V}$. Then, for any source codeword $\mathbf{u} \in \mathcal{V}$ (selected by the information to be transmitted), the encoder transmits

$$\mathbf{X} = (\mathbf{u} - \alpha\mathbf{S} - \mathbf{D}) \bmod \Lambda,$$

where $\mathbf{S}$ is the side-information sequence, and $\mathbf{D}$ is the random dither sequence uniformly distributed over $\mathbf{V}$ and is shared by the encoder and the decoder. At the decoder, the received sequence is given as

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}.$$

The decoder first scales $\mathbf{Y}$ by α and adds the shared dither before modulo reducing it with respect to the quantization lattice. The output is then given as [32, 35]

$$\begin{aligned} \mathbf{Y}' &= (\alpha\mathbf{Y} + \mathbf{D}) \bmod \Lambda \\ &= (\alpha\mathbf{X} + \alpha\mathbf{S} + \alpha\mathbf{Z} + \mathbf{u} - \mathbf{u} + \mathbf{D}) \bmod \Lambda \\ &= (\mathbf{u} - (1-\alpha)\mathbf{X} + \mathbf{Z}) \bmod \Lambda \end{aligned}$$

Because of the dither, $\mathbf{X}$ is uniformly distributed over $\mathbf{V}$ and therefore has the same distribution as $\mathbf{D}$ [32]. Thus, $\mathbf{Y}'$ is equivalent in distribution to [32]

$$\mathbf{Y}' = (\mathbf{u} + \mathbf{Z}') \bmod \Lambda,$$

where $\mathbf{Z}' \equiv ((1-\alpha)\mathbf{D} + \alpha\mathbf{Z}) \bmod \Lambda$. It was shown in [35], that the capacity limit for such a modulo additive noise channel is given as

$$C^* = \frac{1}{2}\log\left(1 + \frac{P_t}{P_Z}\right) - \frac{1}{2}\log\left(2\pi e G(\Lambda)\right), \tag{2.8}$$

where $G(\Lambda)$ is the normalized second moment of the quantizer lattice Λ. $G(\Lambda)$ is upper

bounded by $\frac{1}{12}$ for a uniform quantizer whose Voronoi region is a high dimensional cuboid, and asymptotically approaches $\frac{1}{2\pi e}$ with the dimensionality of Λ going to infinity for a quantizer lattice whose Voronoi region is a high dimensional sphere [39]. We can see that with a lattice that achieves the lowest normalized second moment (ideal quantizer), the capacity limit of the nested lattice DPC scheme is equivalent to Costa's capacity in (2.7). This necessitates the use of a strong source code, along with a capacity achieving channel code in order to get close to Costa's capacity limit.

CHAPTER III

THE GAUSSIAN RELAY CHANNEL

A. Introduction

The relay channel, introduced by van der Meulen in [6], consists of three terminals: the source, the relay, and the destination. The source broadcasts its message to both the relay and the destination. The relay processes the message it receives from the source and forwards the processed signal to the destination, which reconstructs the original message by decoding the signals received from both the source and the relay. The task of the relay is thus to facilitate joint decoding at the destination by means of spatial/temporal diversity.

The capacity of the general relay channel is still not known. Cover and El Gamal [40] derived the tightest upper and lower bounds of the general relay channel using random coding and converse arguments. These two bounds coincide only in few special cases (e.g., the degraded Gaussian relay channel). Since optimum processing at the relay is unknown, several random coding schemes (see, for example, [41, 42, 40, 43, 44, 1, 2]) have been proposed to obtain the lower bound on the capacity. In general, these coding schemes can be divided into two classes: *decode-forward* (DF) and *compress-forward* (CF) [40]. In DF, the relay attempts to decode the transmission from the source before forwarding it to the destination. As a result, the source to relay channel quality acts as the bottle-neck for the overall achievable rate. In order to alleviate the effect of this limitation, CF has been proposed, where the relay does not attempt to decode the signal from the source. Instead, it relies on Wyner-Ziv (WZ) coding [16] (or source coding with side information at the decoder). Since the signals received at the receiver and the destination are different noisy versions of the

same source signal, they are correlated. The relay can exploit this correlation to compress the relay received signal without the knowledge of the received signal at the destination, i.e., the side information; thus it does not attempt to decode. At the destination, the signal is recovered via joint decoding with the presence of side information. CF is desirable over DF when the signals received at the relay and the destination are highly correlated, e.g., when the relay is close to the destination. In addition, as opposed to DF, CF always outperforms direct transmission. Thus, in CF, the relay always helps the source, even if the link between the source and the relay is poor. Therefore, CF gives many rate points that are not achievable with any other coding strategy. More importantly, it was shown in [3] that WZ coding-based CF relaying is optimal in terms of asymptotically achieving the upper bound for receiver cooperation in ad hoc networks.

There are two operating modes in relaying: full-duplex and half-duplex. If the relay is able to transmit and receive signals simultaneously, then we say that it works in the full-duplex mode; otherwise, it works in the half-duplex mode. In this latter mode, the relay either works in a time-division (in which the relay receives and transmits signals in different time slots) or frequency-division manner (in which the relay receives and transmits at different frequencies). Half-duplex is a simpler and cheaper approach [45] because the microwave design challenge (e.g., due to the large difference in the transmitting and receiving signal power levels) associated with the full-duplex mode can be avoided. Therefore, we will focus on half-duplex relaying in this dissertation. Since time- and frequency-division are equivalent from an information-theoretic viewpoint, we will assume time-division multiplexing in the sequel.

Motivated by the wide application of cooperative communications [1, 2, 3, 4, 5] and the information-theoretic importance of the relay channel, several research groups have recently proposed practical cooperative code designs. However, most of them

employ DF. For example, practical user cooperative scheme was given in [44] using space-time coding. Rate-compatible punctured convolutional codes were employed in [4]. A DF scheme based on incremental redundancy designed in [5] exploited optimized convolutional codes in a Rayleigh slow fading environment; it was further shown that this scheme achieved full diversity. A practical turbo-based code design for DF, called distributed turbo coding, was proposed in [46]. A similar code design was also given in [47], and a more advanced turbo-based code design was proposed in [48] for both single-input single-output and multiple-input multiple-output relay system, showing a gap typically 1.0-1.5 dB away from the theoretical limits. Extension of [48] to half-duplex relays was done in [49]. As opposed to the turbo-based designs mentioned above, [50] focused on low-density parity-check (LDPC) code design for the half-duplex DF relaying, reporting a similar loss of 1.2 dB to the theoretical limit. Two more recent works on DF relaying using LDPC codes were documented in [51, 52].

On the other hand, only a few works on practical CF code designs have been published in the literature. This is due to the fact that, despite the significance of WZ coding in CF relaying, practical WZ code design was not well understood until very recently [28, 53, 29]. The first CF design for the half-duplex Gaussian relay channel using WZ coding was devised in our preliminary work of [54, 55], which was soon followed by the publication of [56, 57]. However, no WZ compression or distributed joint source-channel coding (DJSCC) was employed at the relay in [56, 57]; in addition, the source was not allowed to transmit during the relay-transmit period, making the decoder design much easier than in [54, 55] and the overall relay scheme highly suboptimal. A related work by the same authors of [56, 57] was presented in [56], which was based on WZ coding at the relay and used scalar quantization and convolutional codes, but it did not exploit limit-approaching CF designs or give the-

oretical bounds with performance comparisons (to the bounds). A quantize-forward scheme was put forth in [58], which did not exploit WZ coding at the relay. Another quantize-forward schemes was described in [59]. Finally, a comprehensive survey of coding techniques for cooperative communications was given in [60].

Inspired by recent theoretical studies (e.g., [61]) on capacity bounds for the wireless relay channel, we address CF coding for the half-duplex *Gaussian* relay channel − extension of this work to the fading case will be discussed in the next chapter. Guassian (or unconstrained) modulation was assumed in [61], but the complexity of implementing approximate Gaussian modulation (e.g., via turbo trellis-coded modulation and shaping) is high, we thus treat the simplest setup with binary phase shift keying (BPSK) modulation in this work as a first step. However, with BPSK, the signal to be compressed by WZ coding at the relay and the side information at the destination are not jointly Gaussian (as in [61]); in fact, both the source and the side information are Gaussian mixtures, for which the theoretical limit of WZ coding is not known. Consequently, we do not attempt to find the capacity of CF relaying with BPSK modulation.

In this chapter, we first derive the achievable rate of CF coding with BPSK for the half-duplex Gaussian relay channel by considering Slepian-Wolf coded nested scalar quantization (SWCNSQ) [28] as a practical form of WZ coding at the relay. The resulting achievable rate serves as the theoretical bound of our subsequent code design, in which we employ a low-density parity-check (LDPC) code for error protection at the source, and SWCNSQ, i.e., nested scalar quantization followed by Slepian-Wolf (SW) coding [15], at the relay. Since the SW coded bitstream at the relay has to be transmitted over the noisy channel between the relay and the destination, additional error protection is needed. To provide both SW compression and error protection, we resort to distributed joint source-channel coding (DJSCC) [62] of the

quantization indices using a single irregular-repeat accumulate (IRA) code [63]. Once the quantization indices are recovered after IRA decoding, the destination effectively receives the source transmission over two parallel channels. Thus the design process for the LDPC code is different from conventional LDPC code design for point-to-point communication since it has to cater for these two channels. We formulate an optimization problem for designing the degree distributions of the LDPC code using extrinsic information transfer (EXIT) charts [64] and the Gaussian assumption [65], which can be solved using linear programming. In addition, the IRA code is designed using a technique similar to the one in [62] for DJSCC. Utilizing discretized density evolution (DDE) [66], we observe that with asymptotically large block lengths, our code design with optimized LDPC and IRA codes performs $0.11-$ 0.21 dB away from the achievable rate bound of SWCNSQ based CF relaying. Simulations with LDPC/IRA codes of length 2×10^5 bits show a performance loss of $0.27-$ 0.38 dB from the theoretical limit.

The rest of the chapter is organized as follows. The channel model and the WZ coding-based CF coding scheme are described in Section B. We give the achievable rates of CF coding in Section C by employing SWCNSQ as a practical means of WZ coding. We also present several simplifications to the scheme (with negligible performance loss) to facilitate practical implementation. Practical CF relaying using LDPC and IRA codes is described in Section D, followed by the simulation results in Section E. Finally, Section F provides a summary of the chapter.

B. Channel Model and WZ Coding-based CF Relaying

The three-node relay channel is shown in Fig. 11, where d_{sd}, d_{sr} and d_{rd} denote the source to destination, source to relay, and relay to destination distances, respec-

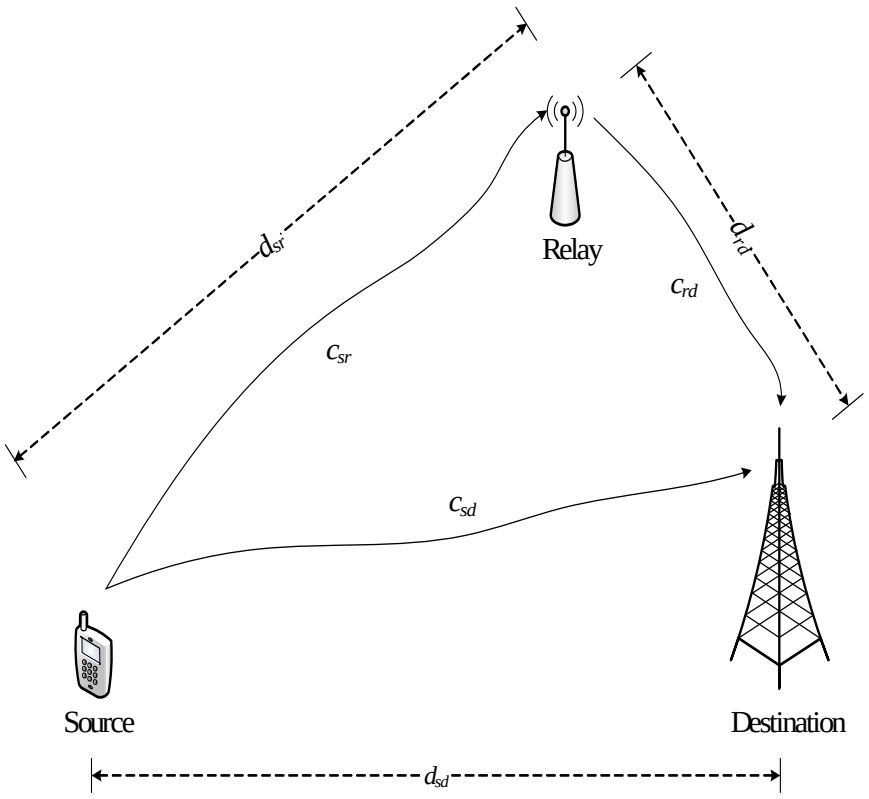


Fig. 11. The relay channel with three nodes: the source, the relay, and the destination.

tively, and c_{sd}, c_{sr}, and c_{rd} are the corresponding channel gains. All links experience additive white Gaussian noises, which are assumed to be of unit variances. As in [46], we assume that the channel coefficients are given by $c_{sr} = c_{sd}\left(\frac{d_{sd}}{d_{sr}}\right)^{l/2}$ and $c_{rd} = c_{sd}\left(\frac{d_{sd}}{d_{rd}}\right)^{l/2}$, where l is the path loss coefficient. Throughout the paper, we fix the path loss coefficient l to three and, without loss of generality, assume that the source to destination channel coefficient is normalized to one, i.e., $c_{sd} = 1$, and that the distance between the source and the destination is also fixed. The exact value of this distance is not important since all distances can be scaled appropriately to obtain the same channel statistics. However, for the sake of expositional clarity, we

assume that the source-to-destination distance is fixed at $d_{sd} = 10$.

In the following, we explain the coding strategy for the half-duplex CF relaying, depicted in Fig. 12.

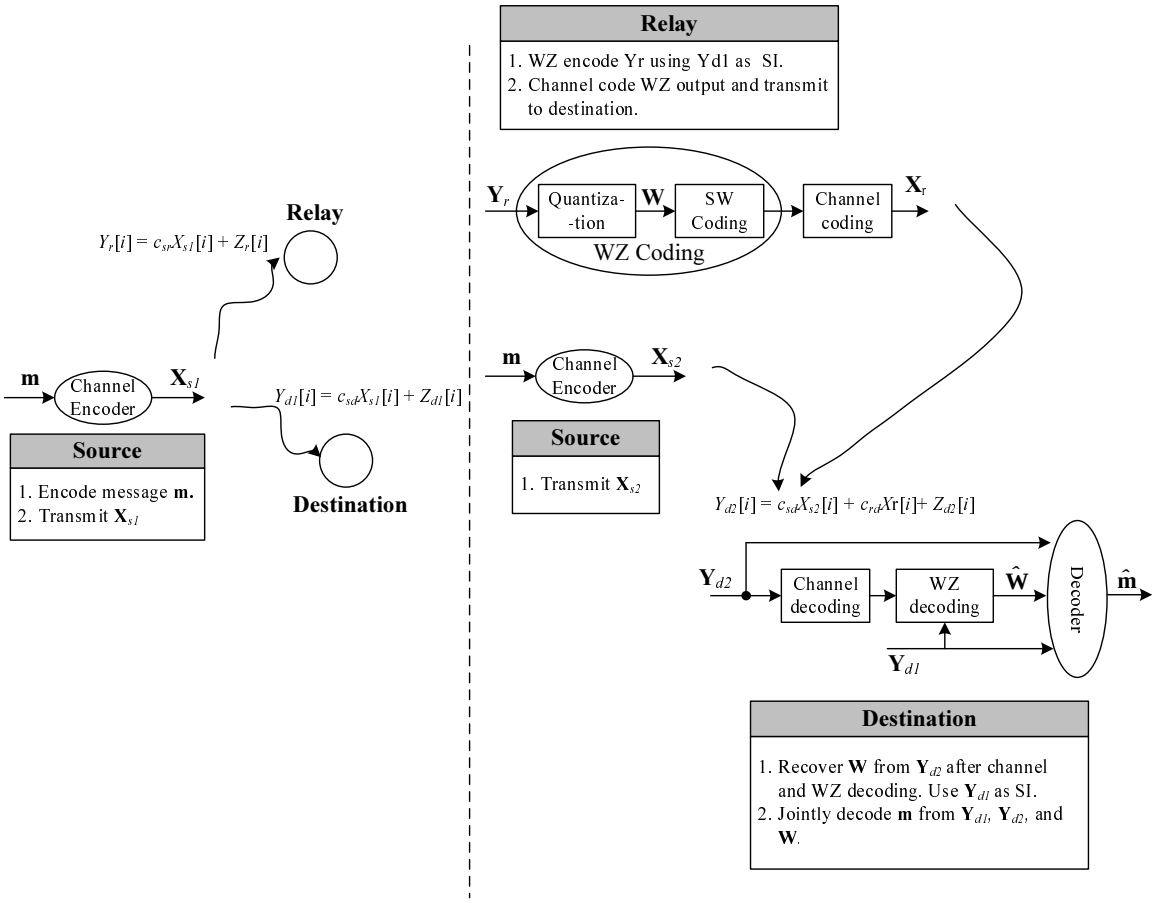


Fig. 12. An illustration of the WZ coding based CF relaying for the half-duplex Gaussian relay channel.

Encoding at the source: A message $\mathbf{m}$ of nR bits at the source is encoded into a length n codeword $\mathbf{X}(\mathbf{m})$. The codeword is split into two non-overlapping parts: $\mathbf{X}_{s1}(\mathbf{m})$ of length αn as the first block of transmission and $\mathbf{X}_{s2}(\mathbf{m})$ of length $\overline{\alpha} n$ as the second block of transmission, where $\alpha \in (0, 1)$ and $\overline{\alpha} = 1 - \alpha$. The two parts of the codeword satisfy the power constraints $\frac{1}{\alpha n} \sum_{i=1}^{\alpha n} X_{s1}[i]^2 \leq P_{s1}$ and

$\frac{1}{\overline{\alpha}n}\sum_{i=1}^{\overline{\alpha}n} X_{s2}[i]^2 \leq P_{s2}$, where P_{s1} and P_{s2} are the source powers during the first and second block of transmission, respectively. We point that that in describing the encoding process at the source, we have deviated slightly from the CF relaying scheme mentioned in [61], where instead of the codeword, the message $\mathbf{m}$ is split into two non-overlapping parts $\mathbf{m}_1$ and $\mathbf{m}_2$. These two messages are then encoded to two independent codewords $\mathbf{X}_{s1}(\mathbf{m}_1)$ and $\mathbf{X}_{s2}(\mathbf{m}_2)$ of lengths αn and $\overline{\alpha}n$, respectively. We will refer to the scheme of [61] as the *message-splitting* (MS) scheme since the original message is split into two parts. On the other hand, the coding strategy in which the codeword instead of the message is split, will be referred to as the *codeword-splitting* (CS) scheme. Information-theoretically, the two schemes result in the same achievable rate. However, for reasons that will be explained in Section D, practical implementation of an MS scheme using LDPC and IRA codes performs worse than that of a CS scheme. Thus we use the latter in our practical CF system design. For this reason, our description is specific to the CS scheme only.

At the frame level, the time interval T for each communication cycle is divided into the relay-receive period $T_1 = \alpha T$ and the relay-transmit period $T_2 = \overline{\alpha}T$ with $T_1 + T_2 = T$.

Relay-receive period: During the relay-receive period, the source transmits $\mathbf{X}_{s1}(\mathbf{m})$ and the signals received at the relay and the destination at time index $i = 1, \ldots, \alpha n$ are given by

$$Y_r[i] = c_{sr}X_{s1}(\mathbf{m})[i] + Z_r[i], \tag{3.1}$$

$$Y_{d1}[i] = c_{sd}X_{s1}(\mathbf{m})[i] + Z_{d1}[i], \tag{3.2}$$

respectively, where Z_r and Z_{d1} are i.i.d. white Gaussian noises with unit power, independent of the source and each other. We thus have a *broadcast channel* in the

relay-receive period.

Relay-transmit period: Note that during T_1, the relay receives a correlated version of the signal received at the destination, because both Y_r and Y_{d1} are noisy replicas of the same signal X_{s1}. Thus during T_2, the relay exploits this correlation to compress $\mathbf{Y}_r$ using WZ coding [16] at a rate of R_{WZ}, using an auxiliary random variable W such that $Y_{d1} \to Y_r \to W$ forms a Markov chain. As mentioned in Chapter II, W in practice can be thought of as the quantized version of Y_r. The WZ coding output is then mapped to a codeword $\mathbf{X}_r(\mathbf{W})$ and transmitted to the destination. The codeword is of length $\overline{\alpha}n$ and satisfies the power constraint $\frac{1}{n}\sum_{i=1}^{\overline{\alpha}n} X_r[i]^2 \leq P_r$, where P_r is the average relay modulation power. At the same time, the source transmits $\mathbf{X}_{s2}(\mathbf{m})$ to the destination. The signal received at the destination is thus

$$Y_{d2}[i] = c_{rd}X_r(\mathbf{W})[i] + c_{sd}X_{s2}(\mathbf{m})[i] + Z_{d2}[i], \quad i = 1, \ldots, \overline{\alpha}n, \tag{3.3}$$

where Z_{d2} is again i.i.d. white Gaussian noise with unit power, independent of the source and the relay signals. We thus have a multiple-access channel (MAC) during the relay-transmit period.

Decoding at the destination: At the destination, an estimate $\hat{\mathbf{W}}$ of the quantization indices is first obtained from $\mathbf{Y}_{d2}$ using $\mathbf{Y}_{d1}$ as decoder side information. The effective transmission channel from source to destination after $\hat{\mathbf{W}}$ has been recovered is shown in Fig. 13, where the destination receives a noisy version of the codeword $\mathbf{X}_s(\mathbf{m})$ over two parallel sub-channels. The length-αn output of sub-channel 1 corresponding to the relay-receive period is $\mathbf{Y}_{d1}$ and $\hat{\mathbf{W}}$, and that of sub-channel 2 corresponding to the relay-transmit period is $\mathbf{Y}_{d2}$, which is of length $\overline{\alpha}n$. The decoder therefore recovers the message $\mathbf{m}$ jointly from $\mathbf{Y}_{d1}$, $\hat{\mathbf{W}}$, and $\mathbf{Y}_{d2}$.

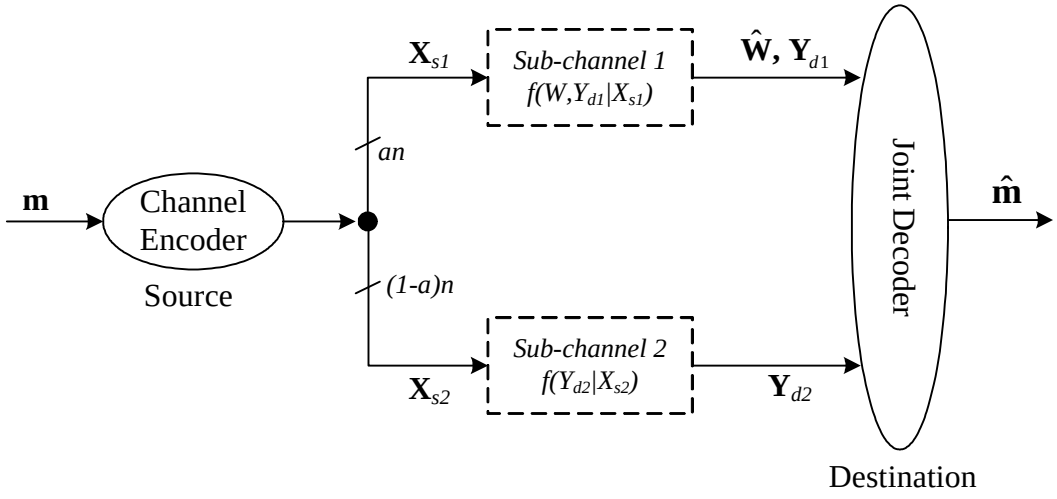


Fig. 13. After decoding the quantization indices $\mathbf{W}$, the destination effectively receives the source transmission over two parallel sub-channels.

C. Achievable Rates of CF Coding Using SWCNSQ with BPSK

In this section, we describe a special case of the WZ coding-based CF relaying scheme of Section B by limiting W to be the output of nested scalar quantization [28] of Y_r. We are interested in this case because our practical CF code design of Section D quantizes Y_r using a nested scalar quantizer before SW coding. In order to gauge the performance of our BPSK modulated code design, we compute the achievable rate R for this special case (with all channel inputs X_{s1}, X_{s2}, and X_r being BPSK modulated). We start with preliminaries and background on SWCNSQ.

1. Preliminaries

The upper bound on the capacity and the achievable rates for CF and DF of the relay channel with the inputs being Gaussian modulated are given in [61]. In that

case, the capacity bounds are relatively simple to derive, since the capacity for the Additive White Gaussian Noise (AWGN) channel and the rate-distortion function of quadratic Gaussian WZ coding exist in close-form expressions. As opposed to the Gaussian input case, with BPSK modulation we have a binary-input AWGN (BIAWGN) channel between each pair of nodes, the capacity of which can only be computed numerically as [67]

$$C^{BIAWGN}(\mathrm{s}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{2\pi}} \log(1 + e^{-2\sqrt{\mathrm{s}}\tau - 2\mathrm{s}}) d\tau, \tag{3.4}$$

where s is the channel signal-to-noise ratio (SNR). We will also come across another type of channel, where in addition to the Gaussian noise, there is an equi-probable binary interference. We call this channel binary-input mixture Gaussian noise (BIMGN) channel. For example, the channel represented by (3.3) becomes a BIMGN channel if X_r is treated as binary input signal and X_{s2} as binary interference. The capacity of such a BIMGNC can be derived as (see Appendix A)

$$C^{BIMGN}(P, P_i, \sigma^2) = 1 - \int_{-\infty}^{\infty} p^+(y) \log\left(1 + \frac{p^-(y)}{p^+(y)}\right) dy, \tag{3.5}$$

with

$$\begin{aligned} p^+(y) &= \frac{1}{2\sqrt{2\pi}\sigma} e^{-(y-\sqrt{P}-\sqrt{P_i})^2/2\sigma^2} + \frac{1}{2\sqrt{2\pi}\sigma} e^{-(y-\sqrt{P}+\sqrt{P_i})^2/2\sigma^2}, \\ p^-(y) &= \frac{1}{2\sqrt{2\pi}\sigma} e^{-(y+\sqrt{P}-\sqrt{P_i})^2/2\sigma^2} + \frac{1}{2\sqrt{2\pi}\sigma} e^{-(y+\sqrt{P}+\sqrt{P_i})^2/2\sigma^2}, \end{aligned}$$

where P and P_i are the received signal and interference powers, respectively, and σ^2 is the AWGN variance. Note that (3.5) degenerates to (3.4) (with $\mathrm{s} = \frac{P}{\sigma^2}$) when $P_i = 0$. In the following, we will abuse the notation by using $C(\cdot)$ and $C(\cdot, \cdot, \cdot)$ to indicate capacity of a BIAWGN and BIMGN channel, respectively.

2. SWCNSQ

As mentioned in Chapter II, WZ coding can be implemented by first quantizing the information sequence followed by SW coding. A framework based on Slepian-Wolf coded nested lattice quantization, i.e., nested lattice quantization of the source, followed by SW coding [15], was put forth for practical WZ coding in [28]. A nested lattice consists of a fine lattice and a coarse lattice that are nested in the sense that any point of the coarse lattice is also a point of the fine lattice, but not vice versa. Nested lattice quantization involves computing the difference between the quantized versions of the source with respect to the fine and coarse lattices. SW coding aims to exploit the remaining correlation between the quantizer output and the decoder side information.

SWCNSQ is the simplest form of Slepian-Wolf coded nested lattice quantization where a 1-D nested scalar lattice is employed for quantization. Fig. 14 depicts a nested scalar quantizer with stepsize q and nesting ratio $M = 4$. When the nesting ratio M goes to infinity, a nested scalar quantizer becomes a conventional uniform scalar quantizer with stepsize q. For the basic rationale behind Slepian-Wolf coded nested lattice quantization for WZ coding, and its performance under the quadratic Gaussian setup, please refer to Chapter II. Readers seeking a more detailed analysis are referred to [28].

3. CF relaying using SWCNSQ

When SWCNSQ is employed for CF relaying, the relay received signal Y_r is first quantized by a nested scalar quantizer with stepsize q and nesting ratio M. The quantization index W is treated as the WZ auxiliary random variable [16] with $W =$

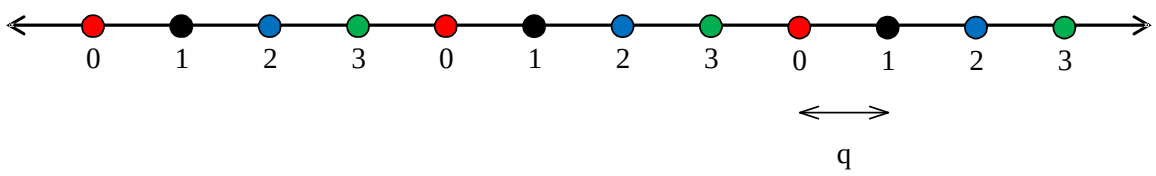


Fig. 14. A nested scalar quantization with stepsize q and nesting ratio $M = 4$.

$w \in \{0, \ldots, M-1\}$ if

$$Y_r \in \bigcup_{k=-\infty}^{\infty} \{x : x \in \mathbb{R}, 0 \leq x - (w + kM)q < q\}.$$

Note that we are using a mid-rise quantizer (with the origin as a quantization cell boundary)[1]. The relay then compresses W to rate R_{WZ} using SW coding. According to [16], R_{WZ} has to satisfy

$$\begin{aligned} R_{WZ} &\geq I(Y_r; W|Y_{d1}) \\ &= H(W|Y_{d1}) - H(W|Y_r, Y_{d1}) \\ &= H(W|Y_{d1}). \end{aligned} \tag{3.6}$$

The last equality (3.6) follows from the fact that $H(W|Y_r, Y_{d1}) = 0$, since W is a function of Y_r. Thus the WZ rate bound reduces to the SW rate for lossless compression of the quantization index W given the decoder side information Y_{d1}. The compressed version of W is then channel coded into X_r at the relay before transmission to the destination during the relay-transmit period.

Recall from Fig. 13 that the destination effectively receives the source trans-

[1]Our numerical results indicate that the achievable rates with a mid-rise quantizer are higher than that of a mid-tread quantizer.

mission over two noisy parallel sub-channels. If the achievable rates on the two sub-channels corresponding to the relay-receive period and relay-transmit period are R_1 and R_2, respectively, an overall rate of

$$R = \alpha R_1 + \overline{\alpha} R_2 \tag{3.7}$$

is then achievable from source to the destination. We now take a close look at rates R_1 and R_2.

During the relay-transmit period, the destination receives transmissions from the source and the relay over a MAC. It can be easily shown that the overall achievable rate for the relay channel is maximized [17] when R_{WZ} and the achievable rate R_2 on sub-channel 2 are on the sum-rate line segment of the MAC capacity region, i.e.,

$$\begin{aligned} R_{WZ} &\leq \frac{\overline{\alpha}}{\alpha}\left[\overline{\beta} C\left(c_{rd}^2 \tilde{P}_r, c_{sd}^2 P_{s2}, 1\right) + \beta C\left(c_{rd}^2 \tilde{P}_r\right)\right] && (3.8) \\ R_2 &\leq \overline{\beta} C\left(c_{sd}^2 P_{s2}\right) + \beta C\left(c_{sd}^2 P_{s2}, c_{rd}^2 \tilde{P}_r, 1\right), && (3.9) \end{aligned}$$

where the normalization factors α and $\overline{\alpha}$ are due to half-duplexing, $\tilde{P}_r = \frac{P_r}{\overline{\alpha}}$, the parameter $\beta \in [0,1]$ indicates the relative position of the operating point on the MAC sum-rate line segment and $\overline{\beta} = 1 - \beta$. For instance, $\beta = 0$ corresponds to one corner point, where the destination first decodes $\mathbf{X}_r(\mathbf{W})$ by treating $\mathbf{X}_{s2}(\mathbf{m})$ as binary interference, and then decodes $\mathbf{X}_s(\mathbf{m})$ after interference cancellation. On the other hand, $\beta = 1$ represents the other corner point which corresponds to the same decoding strategy as $\beta = 0$, except that the roles of $\mathbf{X}_r(\mathbf{W})$ and $\mathbf{X}_{s2}(\mathbf{m})$ are reversed.

Combining the constraints (3.6) and (3.8), we get the following constraint for error free recovery of W

$$H(W|Y_{d1}) \leq \frac{\overline{\alpha}}{\alpha}\left[\overline{\beta} C\left(c_{rd}^2 \tilde{P}_r, c_{sd}^2 P_{s2}, 1\right) + \beta C\left(c_{rd}^2 \tilde{P}_r\right)\right]. \tag{3.10}$$

Since $\mathbf{W}$ along with $\mathbf{Y}_{d1}$ are two noisy versions of $\mathbf{X}_{s1}(\mathbf{m})$, the achievable rate R_1 on sub-channel 1 corresponding to the relay-receive period satisfies

$$\begin{aligned} R_1 &\leq I(W, Y_{d1}; X_{s1}) \\ &= I(Y_{d1}; X_{s1}) + I(W; X_{s1}|Y_{d1}) \\ &= C\left(c_{sd}^2 P_{s1}\right) + C_r, \end{aligned} \tag{3.11}$$

where C_r is the new information rate W conveys about X_{s1} through the relay link (in addition to Y_{d1} through the direct link). This information rate is given as

$$\begin{aligned} C_r &\triangleq I(W; X_{s1}|Y_{d1}) \\ &= H(W|Y_{d1}) - H(W|X_{s1}, Y_{d1}) \\ &= H(W|Y_{d1}) - H(W|X_{s1}), \end{aligned} \tag{3.12}$$

where the last equality follows from the fact that given X_{s1}, the quantization indices are independent of the received signal Y_{d1} at the destination.

Thus the destination can decode the original message at a rate of $R = \alpha R_1 + \overline{\alpha} R_2$ with arbitrarily low probability of error if inequalities (3.9), (3.10) and (3.11) hold. The maximum achievable rate R_{CF} for SWCNSQ is obtained by maximizing R over the feasible set of q's, M's and $0 \leq \alpha, \beta \leq 1$, i.e.,

$$\begin{aligned} R_{CF} &= \max_{0\leq\alpha,\beta\leq 1} \max_{q,M} R \\ &= \max_{0\leq\alpha,\beta\leq 1} \max_{q,M} [\alpha R_1 + \overline{\alpha} R_2] \\ &= \max_{0\leq\alpha,\beta\leq 1} \max_{q,M} \Big[\alpha(C\left(c_{sd}^2 P_{s1}\right) + C_r) + \\ &\qquad \overline{\alpha}\left(\overline{\beta} C\left(c_{sd}^2 P_{s2}\right) + \beta C(c_{sd}^2 P_{s2}, c_{rd}^2 \tilde{P}_r, 1)\right)\Big], \end{aligned} \tag{3.13}$$

while subjecting to the constraint in (3.10).

The optimization in (3.13) can only be performed by a numerical search over the

parameters α, β, q and M. The only term in (3.13) that depends on the quantizer parameters is C_r. In addition, the constraint in (3.10) is also a function of the quantizer. Thus, for fixed α and β, the quantizer design problem in WZ coding at the relay involves finding the quantization step size q and the nesting ratio M that maximizes the rate $C_r = I(W; X_{s1}|Y_{d1})$ such that the constraint (3.10) on $H(W|Y_{d1})$ is satisfied.

Fig. 15 shows C_r as a function of $H(W|Y_{d1})$ for various M (and associated q's). Each curve associated with a specific M is obtained by varying q and recording the corresponding C_r and $H(W|Y_{d1})$. The operational rate curve is identified as the upper concave envelope of the curves corresponding to various M. The optimal C_r for a fixed α and β corresponds to the point on the operational rate curve where (3.10) is satisfied with equality, with the corresponding q and M being the best quantization parameters. Repeating the process over all α and β gives the optimum quantizer. We also include in Fig. 15 the upper bound of C_r, which is $I(X_{s1}; Y_r|Y_{d1})$, achievable when the relay to destination link has infinite capacity, allowing the quantization error between Y_r and W to approach zero.

An important point to note is that the optimization optimization problem mentioned above is different from the usual rate-distortion tradeoff problem in conventional quantizer design, which would have focused on minimizing the mean square distortion between Y_r and its reconstructed version at the destination. Thus a quantizer that minimizes this mean square distortion might not necessarily be the one that maximizes the rate C_r. Indeed, our experiments indicate that for the BPSK modulated relay channel, a vector quantizer with a Gaussian quantization noise achieves a lower C_r than a scalar quantizer even though the former has a smaller mean square distortion associated with it than the latter. This fact is indicated in Fig. 16, in which we consider, besides the NSQ, a Gaussian quantizer with $W = Y_r + Z_q$, where Z_q is a

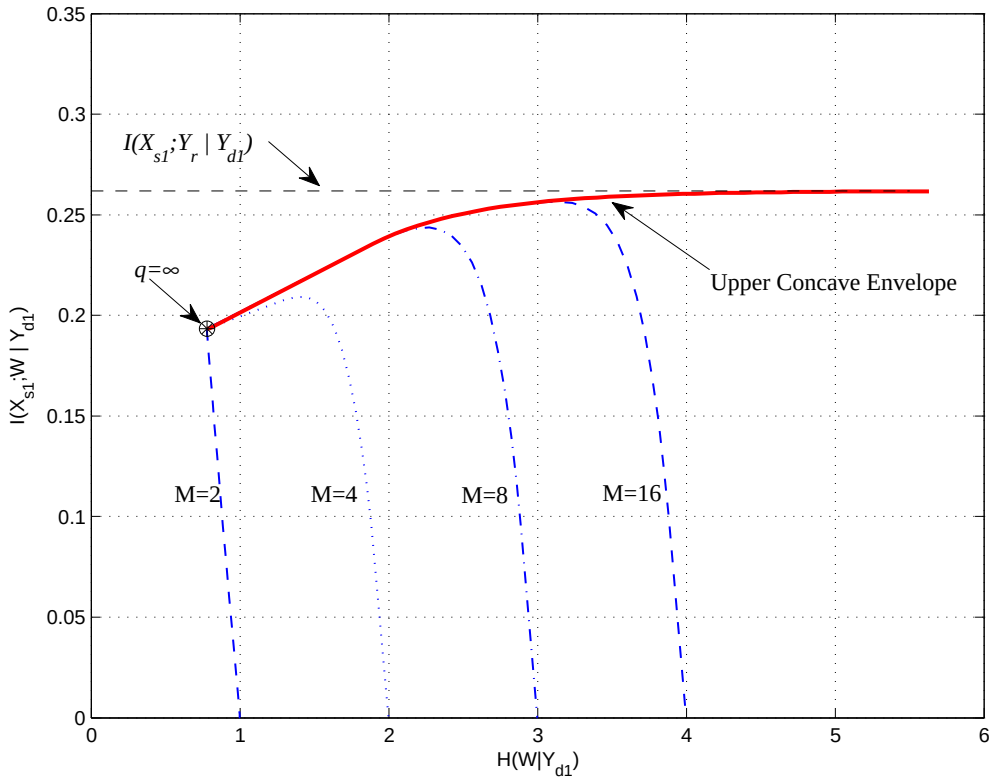


Fig. 15. Rate C_r versus conditional entropy $H(W|Y_{d1})$ curve for $d_{sr} = 9.5$ and $P_{s1} = 0$ dB.

zero-mean Gaussian quantization noise with variance σ_q^2. We obtain the curves in Fig. 16 by varying the quantization noise σ_q^2 over a range of values and plotting the associated mean squared distortion and the achievable rate $I(X_{s1}; W|Y_{d1})$ in Fig. 16(a) and 16(b), respectively. It can be seen that the Gaussian quantizer always achieves a lower mean square distortion than the NSQ. For example, Fig. 16(a) shows that at a WZ compression rate of 0.9 bits, the Gaussian quantizer achieves a mean square distortion which is approximately 1.1 dB better than that of the NSQ. However, this does not guarantee the Gaussian quantizer to outperform the NSQ with respect to

the desired objective, which is the achievable rate $I(X_{s1}; W|Y_{d1})$. Indeed, as shown in Fig. 16(b), at the same WZ compression rate of 0.9 bits, NSQ outperforms the Gaussian quantizer by approximately 0.01 bits.

Note that the achievable rate R_{CF} in (3.13) is given under the transmitting power constraints P_{s1} and P_{s2}. For our practical setup, we assume a constraint P_s on the *average* source transmission power. The power constraints during the relay-receive period and the relay transmit period are related to the average power constraint by $P_{s1} = \frac{\gamma}{\alpha}P_s$, and $P_{s2} = \frac{\overline{\gamma}}{\overline{\alpha}}P_s$, respectively, where $\gamma \in (0, 1)$ determines the power allocation at the transmitter, and $\overline{\gamma} = 1 - \gamma$. Then the achievable rate for an average source power P_s is obtained by maximizing (3.13) over γ.

Next, we briefly discuss numerical computations of the conditional entropies $H(W|Y_{d1})$ and $H(W|X_{s1})$ in C_r, which are required to solve the optimization problem in (3.13).

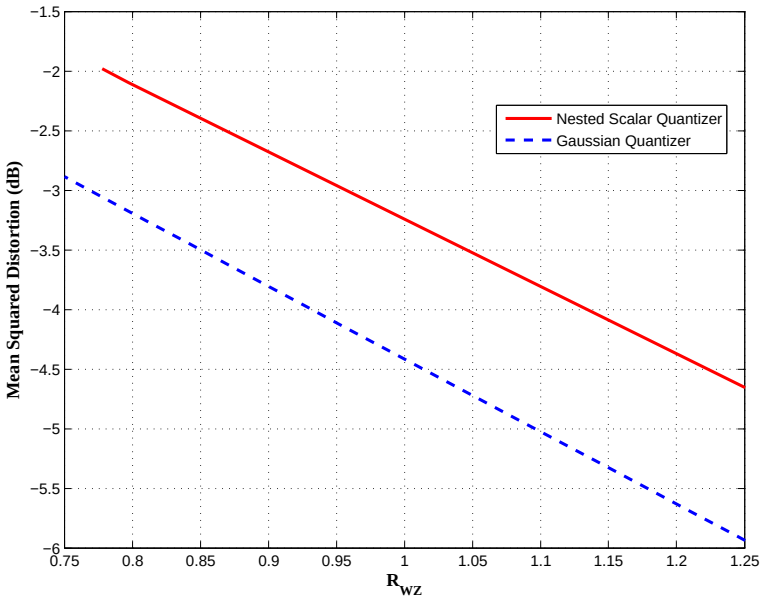
Computation of conditional entropies: In order to compute $H(W|Y_{d1})$ and $H(W|X_{s1})$, we need the conditional probability density of Y_r given Y_{d1}. This can be derived as (see Appendix B)

$$f(y_r|y_{d1}) = \zeta\left(y_{d1}\right) f_g\left(y_r - \tilde{c}_{sr}\right) + \left[1 - \zeta\left(y_{d1}\right)\right] f_g\left(y_r + \tilde{c}_{sr}\right), \tag{3.14}$$

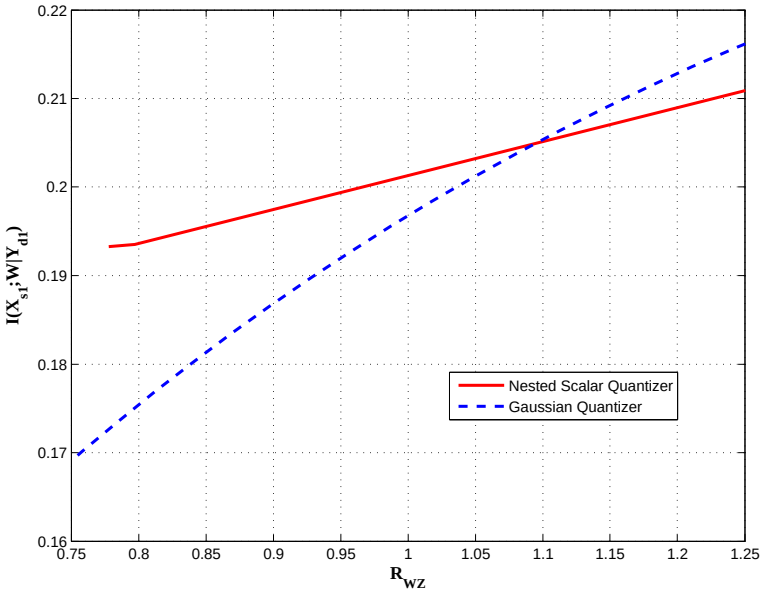
where $\zeta\left(y_{d1}\right) = (1 + e^{-2\tilde{c}_{sd1}y_{d1}})^{-1}$, $\tilde{c}_{sr} = c_{sr}\sqrt{P_{s1}}$, $\tilde{c}_{sd1} = c_{sd}\sqrt{P_{s1}}$ and $f_g(x)$ is the zero mean, unit variance Gaussian probability density function evaluated at x. The conditional probability mass function $P_W(w|y_{d1})$ for a given q and M can be computed as

$$P_W(w|y_{d1}) = \sum_{k=-\infty}^{\infty} \int_{(w+kM)q}^{(w+1+kM)q} f(y_r|y_{d1})dy_r. \tag{3.15}$$

Since $f(y_r|y_{d1})$ in (3.14) is mixture Gaussian, (3.15) can be evaluated using the *erf* functions. In addition, since the Gaussian distribution decays exponentially from the



(a)



(b)

Fig. 16. Comparison of a NSQ with a Gaussian quantizer for $d_{sr} = 9$ and $P_{s1} = 0$ dB. (a) Wyner-Ziv compression rate versus the mean square distortion. (b) Wyner-Ziv compression rate versus the achievable rate $I(X_{s1}; W|Y_{d1})$.

origin, $P_W(w|y_{d1})$ can be well approximated using a finite number of summations with respect to k. Using (3.15), the SW rate can be computed as

$$H(W|Y_{d1}) = -\int f(y_{d1}) \sum_{w=0}^{M-1} P_W(w|y_{d1}) \log\left[P_W(w|y_{d1})\right] dy_{d1}, \tag{3.16}$$

with $f(y_{d1}) = \frac{1}{2}\left(f_g\left(y_{d1} - \tilde{c}_{sd1}\right) + f_g\left(y_{d1} + \tilde{c}_{sd1}\right)\right)$. Since given X_{s1}, Y_r is Gaussian, the conditional entropy $H(W|X_{s1})$ can be computed in a similar manner.

4. Numerical results

In Fig. 17, we plot the achievable rates of CF in (3.13), the DF achievable rate [50], along with the direct transmission rate versus the source-to-relay distance d_{sr} and the relay-to-destination distance d_{rd}. In order to provide a fair comparison, the source power in the direct transmission case is assumed to be equal to the total power $P_s + P_r$ in the relaying case. The average source and relay power is set to $P_s = 0$ dB and $P_r = -12$ dB, respectively. The low power constraint for the relay is motivated by the fact that, in practical cooperative networks, an idle user might agree to relay the transmission of the source provided it allocates only small amounts of transmission resources for relaying. It can be seen from Fig. 17 that CF in general performs better than DF when the relay is close to the destination. In order to provide a better comparison of CF and DF rates, we indicate the difference of the achievable rates of the two schemes as a function of the relay position in Fig. 18. It can be observed that CF outperforms DF for a significant portion of the relay positions, and in fact the closer the relay is to the destination, the more the CF outperforms DF.

Fig. 19 (a) compares R_{CF} with the DF achievable rates as a function of the average source power P_s for a fixed relay position with $d_{sr} = 10$ and $d_{rd} = 2.5$. The average relay power is $P_r = -12$ dB such that the average SNR on the relay to

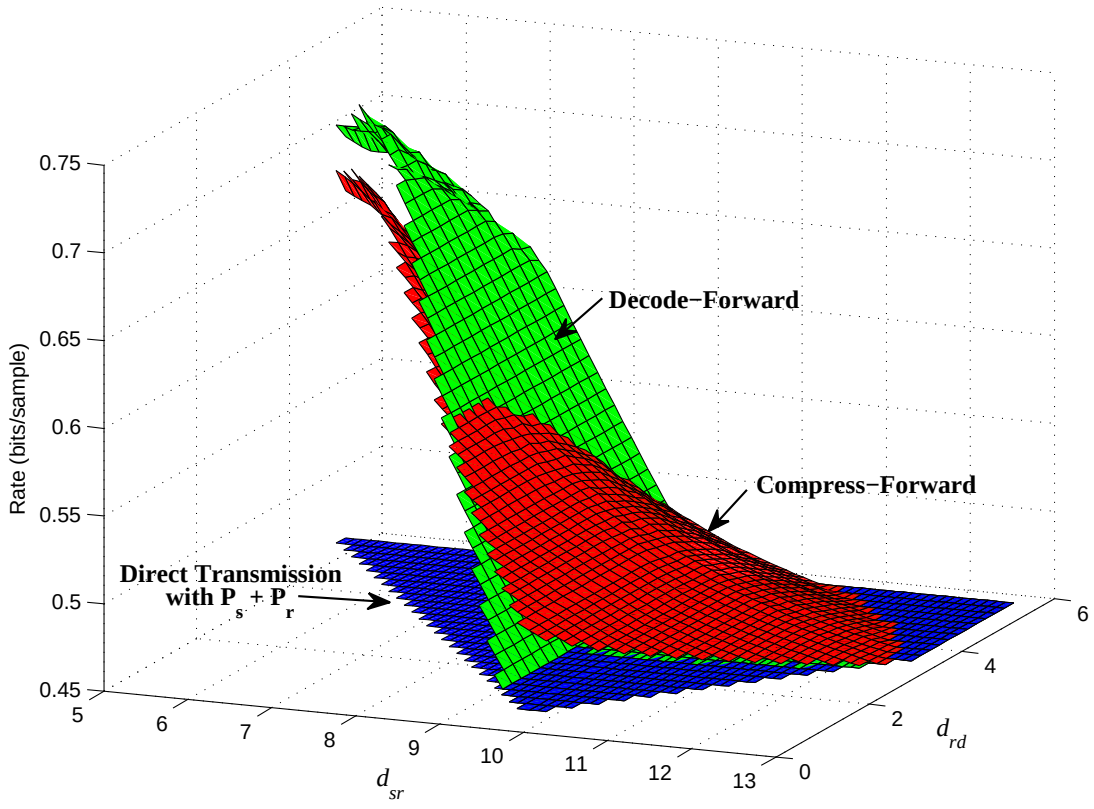


Fig. 17. CF, DF and direct transmission achievable rates as a function of the distances d_{sr} and d_{rd}. The average source and relay power is $P_s = 0$ dB, $P_r = -12$ dB, respectively.

destination link is $c_{rd}^2 P_r = 6$ dB. We also plot the cut-set upper bound for the half-duplex Gaussian relay channel with BPSK modulation, which is obtained by replacing the capacity function of AWGN channels in the upper bound of [61, Proposition 1] by the capacity function (3) of BIAWGN channels. We also plot the achievable rate of CF coding with Gaussian modulation [61], along with the transmission rate when the source transmits directly to the destination without the help of the relay. It is seen that at an overall transmission rate of 0.5 bits/sample (b/s), DF requires

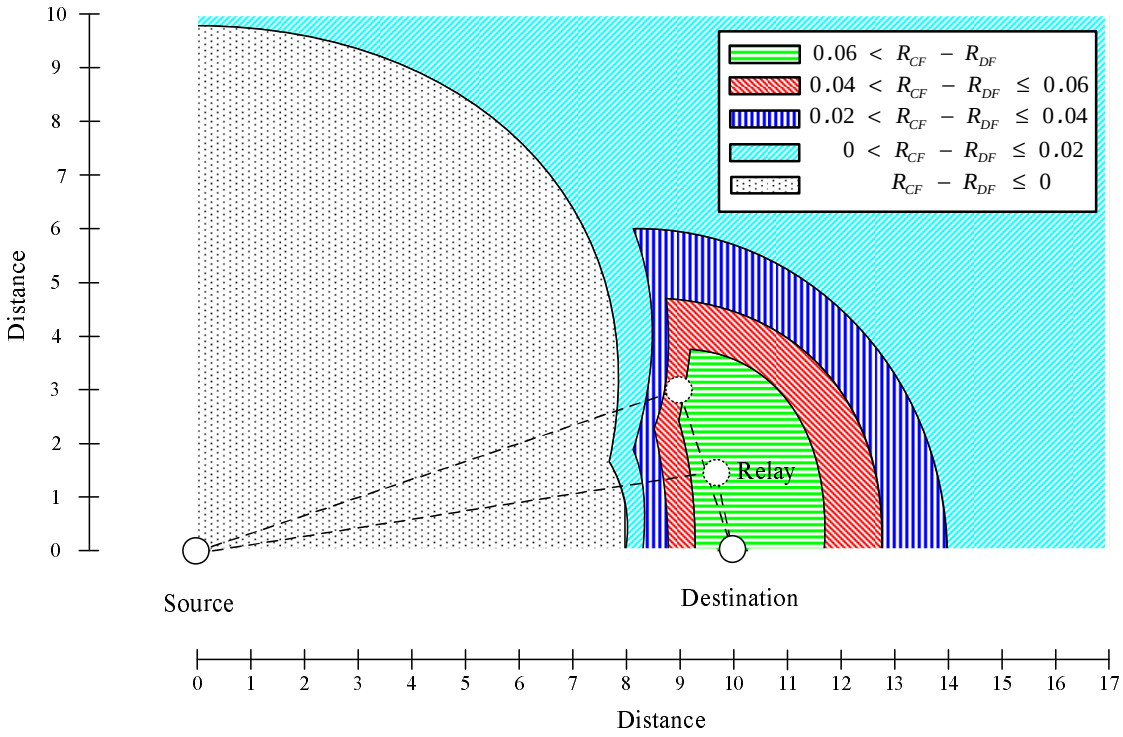
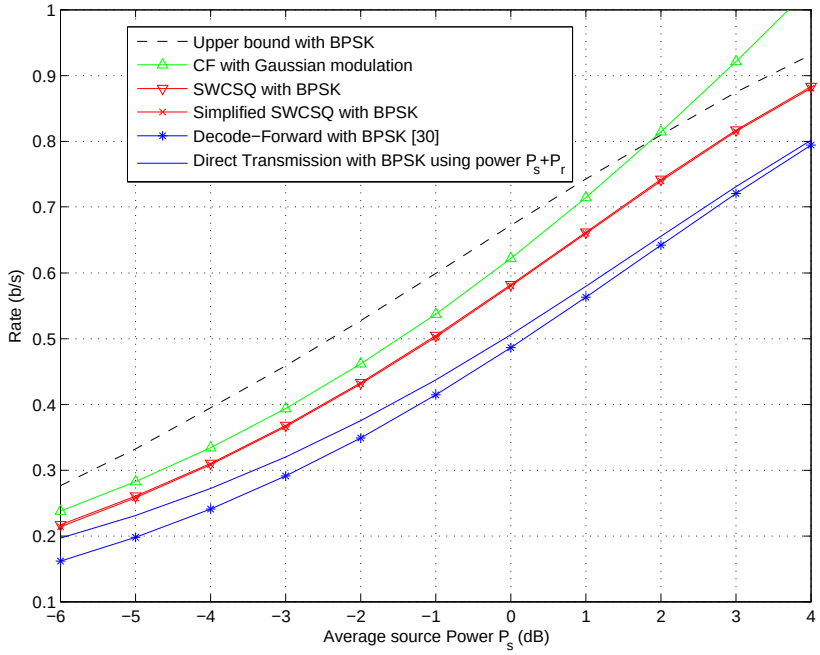
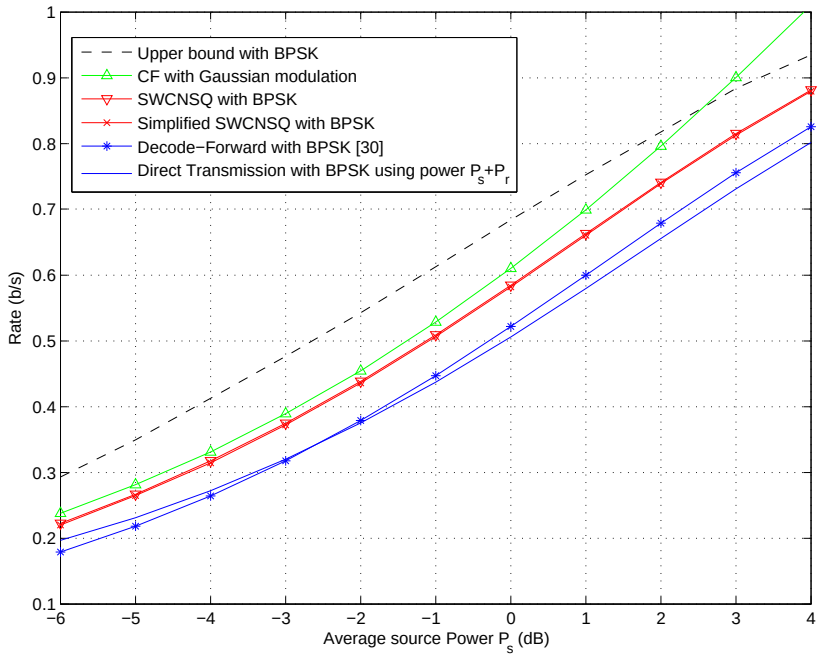


Fig. 18. Comparison of R_{CF} with DF achievable rates as a function of the relay position. The average source and relay power is $P_s = 0$ dB, $P_r = -12$ dB, respectively. The two specific relay positions shown here correspond to the positions for which rate bounds are presented in Fig. 19. Also, it is for the same positions that simulation results of practical CF coding using LDPC and IRA codes are presented in Section E.

1.18 dB more source power than SWCNSQ to achieve the same rate, which in turn outperforms direct transmission by 0.91 dB with source power $P_s + P_r$. Fig. 19 (b) shows the same curves for $d_{sr} = 9.5$ and $d_{rd} = 3.15$ (this setup is chosen such that $c_{rd}^2 P_r$ is 3 dB).



(a)



(b)

Fig. 19. Achievable rates with SWCNSQ as a function of the average source power, with $P_r = -12$ dB and (a) $d_{sr} = 10$, $d_{rd} = 2.5$, and (b) $d_{sr} = 9.5$, $d_{rd} = 3.15$.

5. CF coding using simplified SWCNSQ

In this subsection, we propose a few simplifications to the CF coding scheme described earlier. These simplifications greatly reduce the complexity of a practical implementation, while resulting in very little loss in performance. We list these simplifications along with an *intuitive* understanding of why they result in only a small loss in performance, we also explain how these simplifications help in reducing the complexity of a practical implementation.

- When the MAC time-sharing parameter $\beta \neq 0$, the destination in the relay-transmit period can decode $\mathbf{X}_r(\mathbf{W})$ and $\mathbf{X}_s(\mathbf{m})$ using either joint decoding [68], or rate splitting [69]. These techniques add complexity to practical implementation. Thus, the first simplification we make is to set $\beta = 0$, i.e., we force the destination to first decode $\mathbf{X}_r(\mathbf{W})$ treating $\mathbf{X}_{s2}(\mathbf{m})$ as binary interference. The destination then decodes the source transmission after interference cancellation. This simplifies the design in the sense that it allows decoding of the MAC transmissions using single-user decoders only. For our setup, we assume that the SNR over the relay to destination link is much higher than that of the source to destination link. Thus when $\beta = 0$, i.e., the relay treats the source transmission as interference, the relay is able to transmit at a rate close to the case when there was no interference. Thus both the relay and source can achieve near single-user performance. Because of this, one does not expect a sizable performance loss from this simplication (as confirmed by numerical results).

- Another simplification we make is to set the nested scalar quantization step size to $q = \infty$ and nesting ratio to $M = 2$, thus forcing the quantization index at the relay to be binary. The quantization index is obtained by hard

thresholding the received signal Y_r, i.e., we set $W = 0$ if $Y_r \geq 0$, and $W = 1$ if $Y_r < 0$. One motivation behind using such a quantizer is that the resulting binary quantization indices can be easily SW coded using binary graph-based codes [70, 62]. The main reason why this binary quantization should yield near optimum performance is because the capacity on the relay to the destination link is upper bounded by one, and hence a low rate binary quantizer should suffice. In addition, as mentioned above, the optimum quantizer should maximize the conditional mutual information $I(X_{s1}; W|Y_{d1})$. Since X_{s1} takes on values from a binary alphabet, a binary W should intuitively be able to achieve most of the relaying gain. This is confirmed by Fig. 15 where the difference between $I(X_{s1}; W|Y_{d1})$ when W is binary and the upper bound $I(X_{s1}; Y_r|Y_{d1})$ (when $W = Y_r$ belongs to a continuous alphabet) is small. Hence $M = 2$ and $q = \infty$ should perform well in practice. We point out that good performance was also reported in [56, 57] with a binary quantizer, and leave the problem of quantizer design for the relay channel with higher modulation formats as an interesting topic for future research.

The achievable rates of our simplified SWCNSQ scheme are also included in Fig. 19 alongside the achievable rates of the optimal SWCNSQ scheme, with the former being only 0.004 bits / 0.04 dB worse than the latter. Before moving on to practical code design, we summarize key notations used in Sections B and C in Table I to assist exposition of the material in these sections.

D. Practical CF Code Design

The block diagram of our proposed practical CF relaying system using the simplified scheme of Section C-5 is shown in Fig. 20. We assume that all three nodes have

Table I. List of important variables.

Variable	**Description**	**Comments**
α	Half-duplexing parameter, $\alpha \in (0,1)$	$\overline{\alpha} = 1 - \alpha$
T, T_1, T_2	Length of each tranmission cycle, and relay-receive/transmit period	$T_1 = \alpha T, T_2 = \overline{\alpha} T$
P_s,P_r	Total power constraint at source, relay	$\tilde{P}_r = \frac{P_r}{\alpha}$
X_{s1},X_{s2}	Source transmission during T_1, T_2	
γ	Power allocation parameter at source, $\gamma \in (0,1)$	$\overline{\gamma} = 1 - \gamma$
P_{s1},P_{s2}	Power constraint at source during T_1, T_2	$P_{s1} = \frac{\gamma}{\alpha} P_s, P_{s2} = \frac{\overline{\gamma}}{\overline{\alpha}} P_s$
Y_{d1},Y_{d2}	Destination reception during T_1,T_2	
Y_r	Relay reception during T_1	
W	Quantized version of Y_r	
q	Uniform quantization step size	$q = \infty$ in simplified scheme
M	Nesting ratio	$M = 2$ in simplified scheme
X_r	Relay transmission during T_2	
R_{WZ}	Rate after WZ compression of Y_r at relay	$R_{WZ} \geq H(W\|Y_{d_1})$
C_r	Rate conveyed through the relay about X_{s1}	$C_r = H(W\|Y_{d1}) - H(W\|X_{s1})$
R_1, R_2	Achievable rate on sub-channels 1 and 2 in Fig. 13	$R_1 \leq C(c_{sd}^2 P_{s1}) + C_r$
R	Total transmission rate with CF relaying	$R = \alpha R_1 + \overline{\alpha} R_2$
β	MAC decoding parameter at destination during T_2, $\beta \in [0,1]$	$\beta = 0$ in simplified scheme

perfect knowledge of the channel coefficients, which are used to compute the optimal half-duplexing parameter α and the power allocations P_{s1} and P_{s2}, as explained in Section C-3. In the following, we explain the practical system by describing encoding at the source and the relay, followed by a description of decoding at the destination.

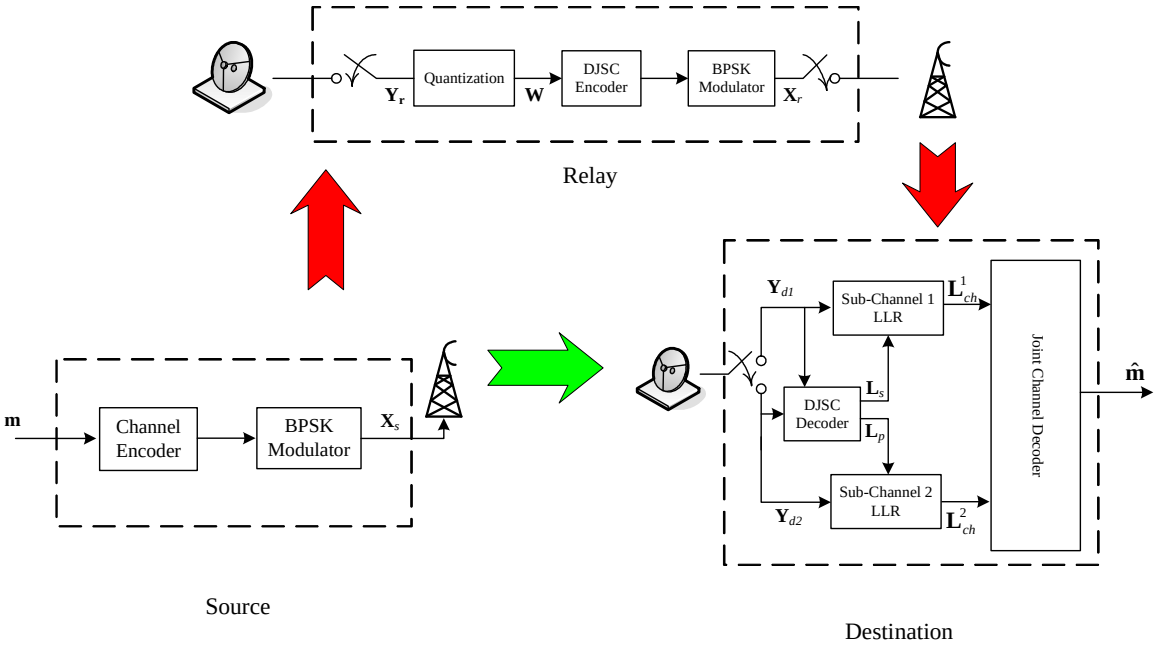


Fig. 20. Block diagram of our proposed CF code design.

1. Encoding at the source

As explained in Section B, in an MS scheme, the source divides its message $\mathbf{m}$ into two non-overlapping parts $\mathbf{m}_1$ and $\mathbf{m}_2$ before encoding them using independent codebooks $\mathcal{C}_1$ and $\mathcal{C}_2$, of rates R_1 and R_2, respectively. It then transmits the first length-αn codeword $\mathbf{X}_{s1}(\mathbf{m}_1)$ during T_1 and the length-$\overline{\alpha}n$ codeword $\mathbf{X}_{s2}(\mathbf{m}_2)$ during T_2. However, for our practical system, instead of MS, we use the CS scheme where the message is encoded using a *single* LDPC code $\mathcal{C}$ of block length n and rate $R = \alpha R_1 + \overline{\alpha} R_2$. During T_1, the first αn codeword bits are mapped to the BPSK constellation points $\pm\sqrt{P_{s1}}$ and broadcast to the relay and the destination. The remaining $\overline{\alpha}n$ codeword bits are BPSK modulated to the constellation points $\pm\sqrt{P_{s2}}$ and transmitted to the destination during T_2. As mentioned earlier, theoretically, the two schemes require

the same source power P_s to achieve a fixed transmission rate. However, as verified by our simulations in Section E, a practical implementation of the CS scheme with a finite length LDPC code requires a lower source power than that of the MS scheme to achieve the same bit error rate (BER) performance. This is because of the following two reasons:

1. It is well known that the BER performance of finite length LDPC codes improves with an increase in block length. Note that since $0 \leq \alpha \leq 1$, the block length of either LDPC code $\mathcal{C}_1$ and $\mathcal{C}_2$ is less than the block length of $\mathcal{C}$. As a result, the CS scheme which uses $\mathcal{C}$ will have a superior BER performance than that of the MS scheme (the BER of MS is the weighted average of the BERs of $\mathcal{C}_1$ and $\mathcal{C}_2$).

2. The second reason why a finite length implementation of CS outperforms that of MS is subtler than the first. As opposed to joint decoding in CS, the destination in MS decodes the two parts of the message, $\mathbf{m}_1$ and $\mathbf{m}_2$, sequentially [61]. After $\mathbf{W}$ has been decoded, the destination cancels the interference from $\mathbf{Y}_{d2}$ and attempts to decode $\mathbf{m}_2$. However, since the LDPC code $\mathcal{C}_2$ suffers a practical coding loss, the power P_{s2} required to achieve a low BER is higher than the theoretical minimum. Thus, in order to obtain a satisfactory performance while decoding $\mathbf{m}_2$, the source has to use a higher transmission power P_{s2}. Consequently, the decoder for $\mathbf{W}$ sees a stronger interference from the source. We assume that the system does not allow an increase in relay transmission power, and therefore, the only way to compensate for the increased interference is by increasing P_{s1} (increasing P_{s1} increases the correlation between Y_{d1} and W and hence facilitates the decoding of W at the destination). Thus, in an MS scheme, P_{s1} is adversely affected indirectly because of an increase in P_{s2}, and directly

because of the practical coding losses of $\mathbf{W}$ and $\mathbf{m}_1$ decoders. In contrast, in a CS scheme, since the message is recovered jointly from $\mathbf{Y}_{d1}$ and $\mathbf{Y}_{d2}$, one can avoid the indirect affect by keeping P_{s2} at the theoretical minimum while increasing P_{s1} to obtain satisfactory BER performance, thus resulting in a lower required source transmission power than MS.

2. DJSCC at the relay

While presenting the achievable rate of CF relaying, Section C assumes that the relay encodes the quantization indices $\mathbf{W}$ using separate SW source coding (for compression) and channel coding (for forward error protection). Since practical SW coding is done using a channel code [70], *separate* source-channel coding at the relay (with side information Y_{d1} at the destination) requires two channel codes: one for SW compression and another for forward error protection. However, just like Shannon's classic separation principle [17], the separation principle [71] for the noisy channel SW/WZ coding problem only holds asymptotically (i.e., with infinite code length). In practical designs with *finite* code length, well-designed *joint* source-channel coding with side information (or DJSCC) outperforms a separate design [72]. Thus, we propose to code the quantization indices $\mathbf{W}$ at the relay using DJSCC. In the following, we explain how DJSCC can be implemented using IRA codes [62].

The basic idea of DJSCC using an IRA code is depicted in Fig. 21. The binary indices $\mathbf{W}$ obtained after hard-thresholding (quantizing) the received sequence $\mathbf{Y}_r$ are encoded using an $(n, \alpha n)$ systematic IRA code. However, the relay transmits only the $\overline{\alpha} n$ parity bits to the destination, which are BPSK modulated to the constellation points $\pm\sqrt{\tilde{P}_r}$. These parity bits are not only used for SW coding, but also for error protection. Note that we have the side information $\mathbf{Y}_{d1}$ available at the destination, which can be viewed as a "noisy" version of the systematic bits $\mathbf{W}$. Thus, even

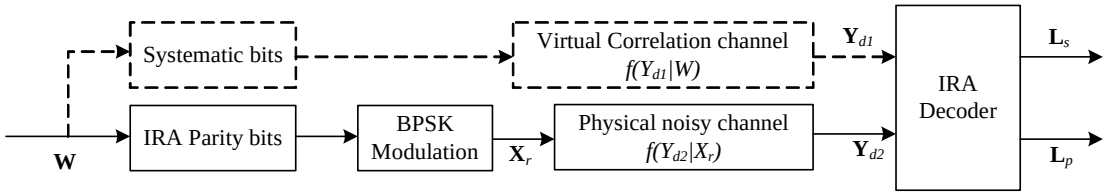


Fig. 21. DJSCC of the quantization indices $\mathbf{W}$ and the side information $\mathbf{Y}_{d1}$ using a systematic IRA code that is designed for both the physical noisy channel and the "virtual" correlation channel between $\mathbf{W}$ and $\mathbf{Y}_{d1}$. The IRA decoder outputs the extrinsic LLRs $\mathbf{L}_s$ and $\mathbf{L}_p$ of the systematic and parity bits, respectively.

though the systematic bits are not transmitted over the physical channel, $\mathbf{Y}_{d1}$ can be viewed as the output of a virtual correlation channel characterized by the probability density function (pdf) $f\,(y_{d1}|W)$. Using the concept of DJSCC, we can examine the constraint on conditional entropy $H(W|Y_{d1})$ which is necessary for successful recovery of $\mathbf{W}$ as follows. The total information to be transmitted from the relay to the destination is $\alpha nH(W)$ bits. The information received at the destination from the virtual correlation channel is $\alpha nI(W;Y_{d1})$ bits, and from the physical noisy channel is $n\overline{\alpha}I(W;Y_{d2})$. The condition necessary for successful decoding of $\mathbf{W}$ is that the information to be transmitted is less than or equal to the information received, i.e.,

$$\begin{aligned} \alpha nH(W) &\leq \overline{\alpha}nI(W;Y_{d2}) + \alpha nI(W;Y_{d1}) \\ &= \overline{\alpha}nI(W;Y_{d2}) + \alpha nH(W) + \alpha nH(W|Y_{d1}) \\ \Rightarrow\ \alpha H(W|Y_{d1}) &\leq \overline{\alpha}I(W;Y_{d2}), \end{aligned}$$

which is the same constraint as in (3.10). This also implies that if the channel code being used for DJSCC is a capacity achieving code on the combined physical noisy channel and the virtual correlation channel, it will also be able to achieve the DJSCC limit of (3.10).

3. Decoding at the destination

The first step in the decoding process is to recover the quantization indices $\mathbf{W}$ by applying iterative sum-product algorithm (SPA) decoding on the IRA code graph. The variable/bit nodes of the IRA graph are divided into two categories: the first category corresponds to the systematic bits $\mathbf{W}$, and the second one to the parity bits $\mathbf{X}_r$. As explained earlier, the side information $\mathbf{Y}_{d1}$ can be thought of as an output of a virtual correlation channel which has the systematic bits as inputs. Thus, the decoder uses $\mathbf{Y}_{d1}$ to calculate the channel log-likelihood ratios (LLR) for the systematic bits. With the quantization step size $q = \infty$, these LLRs can be evaluated using the conditional pdf $f(y_r|y_{d1})$ given in (3.14) as

$$\begin{aligned} L^s_{ch}[i] &= \log \frac{\mathrm{P}\left(W[i]=0 \,\middle|\, y_{d1}[i]\right)}{\mathrm{P}\left(W[i]=1 \,\middle|\, y_{d1}[i]\right)} = \log \frac{f\left(y_{d1}[i] \,\middle|\, W[i]=0\right)}{f\left(y_{d1}[i] \,\middle|\, W[i]=1\right)} \\ &= \log \frac{\int_0^\infty f(y_r|y_{d1}[i])dy_r}{1-\int_0^\infty f(y_r|y_{d1}[i])dy_r} \qquad (3.17) \\ &= \log \frac{1+\left(2\zeta\left(y_{d1}[i]\right)-1\right)\mathrm{erf}\left(\frac{\tilde{c}_{sr}}{\sqrt{2}}\right)}{1-\left(2\zeta\left(y_{d1}[i]\right)-1\right)\mathrm{erf}\left(\frac{\tilde{c}_{sr}}{\sqrt{2}}\right)}, \qquad i=1,\ldots,\alpha n, \qquad (3.18) \end{aligned}$$

where $\zeta\left(y_{d1}\right)$ and $\tilde{c}_{sr}$ are defined in Section C-3, and $\mathrm{erf}\left(x\right) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. In going from (3.17) to (3.18), we note that

$$\begin{aligned} \int_0^\infty f(y_r|y_{d1}[i])dy_r &= \zeta\left(y_{d1}[i]\right)\int_{-\tilde{c}_{sr}}^\infty f_g(y_r)dy_r + \left(1-\zeta\left(y_{d1}[i]\right)\right)\int_{\tilde{c}_{sr}}^\infty f_g(y_r)dy_r \\ &= \zeta\left(y_{d1}[i]\right)\int_{-\infty}^{\tilde{c}_{sr}} f_g(y_r)dy_r \\ &\quad + \left(1-\zeta\left(y_{d1}[i]\right)\right)\left(1-\int_{-\infty}^{\tilde{c}_{sr}} f_g(y_r)dy_r\right) \\ &= 1-\zeta\left(y_{d1}[i]\right)+\left(2\zeta\left(y_{d1}[i]\right)-1\right)\left(\frac{1}{2}+\int_0^{\tilde{c}_{sr}} f_g(y_r)dy_r\right) \\ &= \frac{1}{2}\left(1+\left(2\zeta\left(y_{d1}[i]\right)-1\right)\mathrm{erf}\left(\frac{\tilde{c}_{sr}}{\sqrt{2}}\right)\right), \qquad (3.19) \end{aligned}$$

where the first equality is due to (3.14).

As for the parity bit nodes, the corresponding LLRs are obtained by treating the transmission from the source as binary interference. These LLRs can be evaluated as

$$\begin{aligned} L_{ch}^{p}[i] &= \log \frac{\mathrm{P}\left(X_r[i] = +\sqrt{\tilde{P}_r} \mid y_{d2}[i]\right)}{\mathrm{P}\left(X_r[i] = -\sqrt{\tilde{P}_r} \mid y_{d2}[i]\right)} = \log \frac{f\left(y_{d2}[i] \mid X_r[i] = +\sqrt{\tilde{P}_r}\right)}{\mathrm{P}\left(y_{d2}[i] \mid X_r[i] = +\sqrt{\tilde{P}_r}\right)} \\ &= \log \frac{\sum_{d=\pm 1} f_g\left(y_{d2}[i] + d\tilde{c}_{sd2} - \tilde{c}_{rd}\right)}{\sum_{d=\pm 1} f_g\left(y_{d2}[i] + d\tilde{c}_{sd2} + \tilde{c}_{rd}\right)}, \quad i = 1, \ldots, \overline{\alpha} n, \end{aligned} \tag{3.20}$$

where $\tilde{c}_{sd2} = c_{sd}\sqrt{P_{s2}}$ and $\tilde{c}_{rd} = c_{rd}\sqrt{\tilde{P}_r}$. Using the channel LLRs in (3.18) and (3.20), the destination performs iterative SPA decoding on the IRA graph until some stopping criterion is met[2]. At the end of the iterative process, the decoder obtains extrinsic LLRs $\mathbf{L}_s$ and $\mathbf{L}_p$ of the systematic and parity bits, respectively.

We now look at decoding of the original message from the source. Recall that the message $\mathbf{m}$ in a CS scheme is encoded using a single LDPC code, but the destination effectively sees the length-n codeword transmitted over two parallel sub-channels. Thus the bit nodes of the LDPC decoding graph can be divided into two groups: the first αn nodes corresponding to codeword bits received over sub-channel 1, and the remaining $\overline{\alpha} n$ nodes corresponding to bits received over sub-channel 2. In the following, we refer to the two categories as bit nodes of type-1 and type-2, respectively. According to the information-theoretic scheme of Section C-5, the decoded version of $\mathbf{X}_r$ is used to cancel interference from $\mathbf{Y}_{d2}$, and the resulting interference-free signal is used for decoding the bits received from the source over sub-channel 2. However, for our practical system, instead of using the hard-threshold decoded version of $\mathbf{X}_r$, we use the corresponding extrinsic LLR $\mathbf{L}_p$ to implement a soft interference cancellation strategy. Specifically, if $x = \sqrt{\tilde{P}_r}$, the LLRs to type-2 bit nodes can be evaluated as

[2]In our simulations, we stop iterations when either a valid codeword is decoded, or a maximum number of iterations are reached.

follows.

$$\begin{aligned} L^2_{ch}[i] &= \log \frac{\mathrm{P}\left(X_{s2}[i] = +\sqrt{P_{s2}} \;\middle|\; y_{d2}[i], L_p[i]\right)}{\mathrm{P}\left(X_{s2}[i] = -\sqrt{P_{s2}} \;\middle|\; y_{d2}[i], L_p[i]\right)} = \log \frac{f\left(y_{d2}[i] \;\middle|\; X_{s2}[i] = +\sqrt{P_{s2}}, L_p[i]\right)}{f\left(y_{d2}[i] \;\middle|\; X_{s2}[i] = -\sqrt{P_{s2}}, L_p[i]\right)} \\ &= \log \frac{\sum_{d=\pm x} \mathrm{P}\left(X_r = d \;\middle|\; L_p[i]\right) f_g\left(y_{d2}[i] - \tilde{c}_{sd2} + dc_{rd}\right)}{\sum_{d=\pm x} \mathrm{P}\left(X_r = d \;\middle|\; L_p[i]\right) f_g\left(y_{d2}[i] + \tilde{c}_{sd2} + dc_{rd}\right)} \\ &= \log \frac{e^{L_p[i]} f_g\left(y_{d2}[i] - \tilde{c}_{sd2} + \tilde{c}_{rd}\right) + f_g\left(y_{d2}[i] - \tilde{c}_{sd2} - \tilde{c}_{rd}\right)}{e^{L_p[i]} f_g\left(y_{d2}[i] + \tilde{c}_{sd2} + \tilde{c}_{rd}\right) + f_g\left(y_{d2}[i] + \tilde{c}_{sd2} - \tilde{c}_{rd}\right)}, \end{aligned} \tag{3.21}$$

for $i = 1, \ldots, \overline{\alpha} n$. Note that when $L_p[i] = \pm\infty$, the channel LLR in (3.21) is the same as that obtained using hard interference cancellation. Similarly, instead of using the hard-threshold decoded version of $\mathbf{W}$, we use the corresponding extrinsic output LLR $\mathbf{L}_s$ to obtain the channel LLRs for type-1 bit nodes. These are given by

$$\begin{aligned} L^1_{ch}[i] &= \log \frac{\mathrm{P}\left(X_{s1}[i] = +\sqrt{P_{s1}} \;\middle|\; y_{d1}[i], L_s[i]\right)}{\mathrm{P}\left(X_{s1}[i] = -\sqrt{P_{s1}} \;\middle|\; y_{d1}[i], L_s[i]\right)} \\ &= \log \frac{f\left(y_{d1}[i] \;\middle|\; X_{s1}[i] = +\sqrt{P_{s1}}\right)}{f\left(y_{d1}[i] \;\middle|\; X_{s1}[i] = -\sqrt{P_{s1}}\right)} + \log \frac{\mathrm{P}\left(X_{s1}[i] = +\sqrt{P_{s1}} \;\middle|\; L_s[i]\right)}{\mathrm{P}\left(X_{s1}[i] = -\sqrt{P_{s1}} \;\middle|\; L_s[i]\right)} \\ &= 2\tilde{c}_{sd} y_{d1}[i] + \log \frac{\mathrm{P}\left(X_{s1}[i] = +\sqrt{P_{s1}} \;\middle|\; L_s[i]\right)}{\mathrm{P}\left(X_{s1}[i] = -\sqrt{P_{s1}} \;\middle|\; L_s[i]\right)}. \end{aligned} \tag{3.22}$$

Let us take a look at the second term on the right hand side of (3.22). As shown in Fig. 22, W can be viewed as an output of a binary-symmetric channel (BSC) with X_{s1} as the input, with the crossover probability ϵ equal to the probability that the unit-variance zero-mean AWGN is greater than $\tilde{c}_{sr}$. Thus, we have $\epsilon = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\tilde{c}_{sr}}{\sqrt{2}}\right)\right]$. However, the DJSCC decoder outputs some a-priori information on the quantization indices W in the form of the LLRs $L_s[i]$, $i = 1, \ldots, \alpha n$. Using this a-priori information, the second term on the right hand side of (3.22) can be given as (Using $x = \sqrt{P_{s1}}$ for

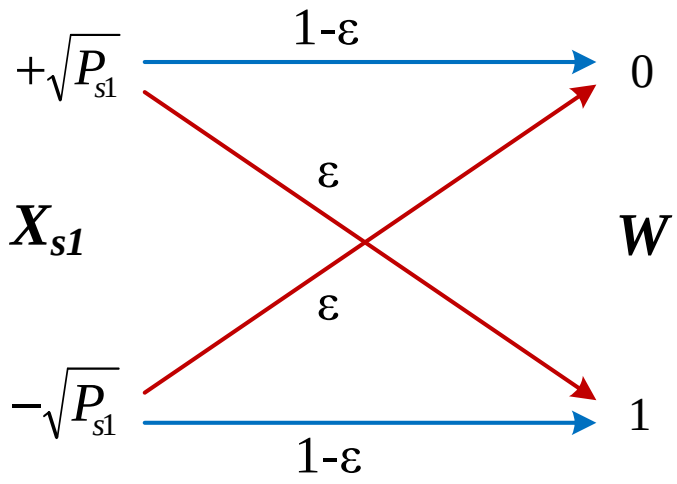


Fig. 22. Equivalent channel from X_{s1} to W. The crossover probability from X_{s1} to W is $\epsilon = \frac{1}{2}\left[1 - \text{erf}\left(\frac{\tilde{c}_{sr}}{\sqrt{2}}\right)\right]$.

notational convenience)

$$\begin{aligned}
\frac{\mathrm{P}\left(X_{s1}[i] = +x \,\middle|\, L_s[i]\right)}{\mathrm{P}\left(X_{s1}[i] = -x \,\middle|\, L_s[i]\right)} &= \frac{\sum_{w=0,1} \mathrm{P}\left(X_{s1}[i] = x, W = w \,\middle|\, L_s[i]\right)}{\sum_{w=0,1} \mathrm{P}\left(X_{s1}[i] = -x, W = w \,\middle|\, L_s[i]\right)} \\
&= \frac{\sum_{w=0,1} \mathrm{P}\left(W = w \,\middle|\, L_s[i]\right) \mathrm{P}\left(X_{s1}[i] = +x \,\middle|\, W = w\right)}{\sum_{w=0,1} \mathrm{P}\left(W = w \,\middle|\, L_s[i]\right) \mathrm{P}\left(X_{s1}[i] = -x \,\middle|\, W = w\right)} \\
&= \frac{\sum_{w=0,1} \mathrm{P}\left(W = w \,\middle|\, L_s[i]\right) \mathrm{P}\left(W = w \,\middle|\, X_{s1}[i] = +x\right)}{\sum_{w=0,1} \mathrm{P}\left(W = w \,\middle|\, L_s[i]\right) \mathrm{P}\left(W = w \,\middle|\, X_{s1}[i] = -x\right)} \\
&= \frac{e^{L_s[i]}(1-\epsilon) + \epsilon}{(1-\epsilon) + e^{L_s[i]}\epsilon}
\end{aligned}$$

Thus, the channel LLR to the type-1 bit nodes are given as

$$L_{ch}^{1}[i] = 2\tilde{c}_{sd} y_{d1}[i] + \log \frac{e^{L_s[i]}(1-\epsilon) + \epsilon}{(1-\epsilon) + e^{L_s[i]}\epsilon}, \quad i = 1, \ldots, \alpha n, \tag{3.23}$$

Using (3.21) and (3.23), the decoder obtains the channel LLR for the two types of bit nodes and performs iterative decoding on the LDPC graph until a stopping criterion is met. Finally, the output LLRs obtained after the iterations are used to obtain an estimate of the LDPC codeword transmitted by the source.

4. Design of degree distributions for asymptotically large block length

The code design requires optimizing the degree distributions of the IRA code being used for DJSCC, as well as the LDPC code being used for the transmission of the original source message. For designing the degree distributions for both codes, we use the Gaussian approximation [65] and EXIT charts strategy [64]. In the following, we discuss the design of the two codes separately.

a. Design of IRA code for DJSCC

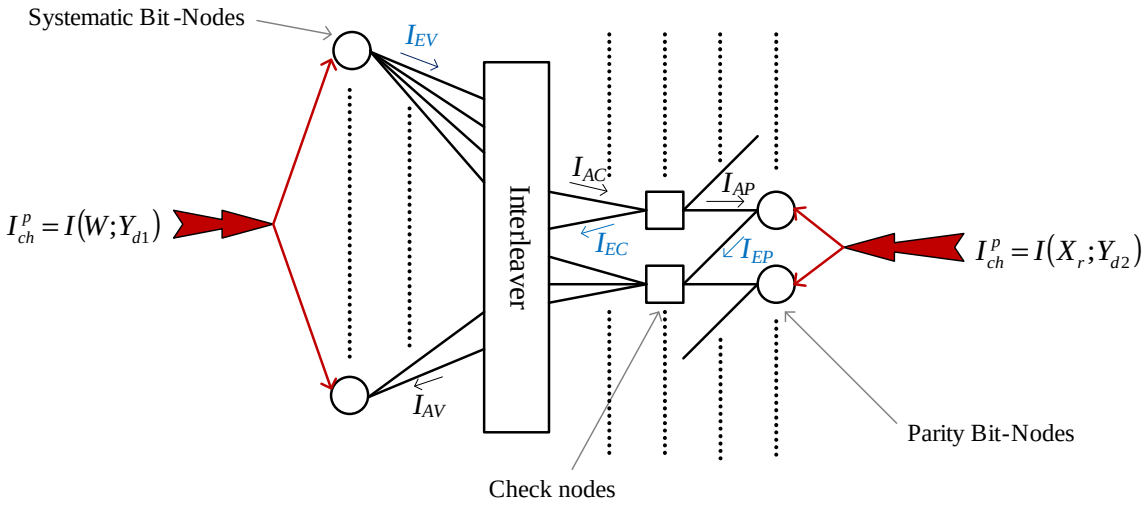


Fig. 23. Information flow in the IRA code decoding graph for DJSCC.

The design methodology we use for designing the degree distributions for the IRA code is similar to the one in [62], except that we use the EXIT chart strategy in

addition to the Gaussian approximation in [62]. For the optimization of the degree distributions, we assume that all LLR messages on the graph edges in the iterative SPA decoding are Gaussian. It can be verified that both the physical noisy channel, as well as the virtual correlation channel are symmetric (see Appendix C), and hence if an LLR at a particular iteration has mean μ, it will have a variance 2μ [73]. Let $J(\mu)$ be the information that the LLR conveys about the bit node that the edge is connected to. Since the channels are symmetric, we can assume the transmission of an all-zero codeword [73]. In addition, since the messages are assumed to be Gaussian, $J(\mu)$ can be related to the capacity of a BIAWGN channel in (3.4) by $J(\mu) = C\left(\frac{\mu}{2}\right)$ [64]. The information flow in the IRA code decoding graph for DJSCC is shown in Fig. 23. The bit nodes on the left hand side of Fig. 23 correspond to the systematic bits (associated with the virtual correlation channel), whereas the right ones to the parity bits (associated with the physical noisy channel). Due to the structure of the IRA decoding graph [63], the right bit nodes always have degree two, with the two edges connected to adjacent check nodes. For our design, we assume that the check node degree distribution is regular, i.e., all check nodes have the same degree d_c. The design parameters therefore are the systematic bit degree distribution coefficients. Let $I_{AC} \in [0, 1]$ be the average a-priori input information to the check nodes. Let I_{AP} be the a-priori information to the parity bit nodes from the check nodes, and I_{EP} be the extrinsic information from the parity bit nodes to the check nodes. In order to calculate the output extrinsic information from the check nodes, we use the approximate bit and check node information duality [74] according to which

$$I_{AP} \approx 1 - J\left(d_c J^{-1}(1 - I_{AC}) + J^{-1}(1 - I_{EP})\right). \tag{3.24}$$

The extrinsic information I_{EP} from the parity bit nodes is then given as

$$I_{EP} = J\left(J^{-1}(I_{AP}) + J^{-1}(I_{ch}^{p})\right), \tag{3.25}$$

where $I_{ch}^{p} = I(X_r; Y_{d2})$ is the input channel information to the parity bit nodes over the physical noisy channel. Then, following the spirit of [62], we compute I_{EP}^{*} for a given I_{AC} which is obtained by substituting (3.24) in (3.25) and solving for I_{EP}. For a given I_{AC}, this is akin to running belief-propagation (BP) iterations on the sub-graph to the right of the check nodes until a fixed point is reached. The extrinsic information from the check nodes to the systematic bit nodes is then given as

$$I_{EC}\left(I_{AC}\right) \approx 1 - J\left((d_c - 1)J^{-1}(1 - I_{AC}) + 2J^{-1}(1 - I_{EP}^{*})\right). \tag{3.26}$$

At the systematic bit nodes, the extrinsic information as a function of I_{AC} is given as [64]

$$I_{EV}\left(I_{AC}\right) = \sum_{i=2}^{D_v} \omega_i J\left(J^{-1}(I_{ch}^{s}) + (i-1)J^{-1}\left(I_{EC}(I_{AC})\right)\right), \tag{3.27}$$

where $I_{ch}^{s} = I(W; Y_{d1})$ is the input channel information the systematic parity bit nodes over the virtual correlation channel, ω_i is the fraction of edges connected to degree-i systematic bit nodes, and D_v is the maximum degree. Then for an error probability zero, we need [64]

$$I_{EV}\left(I_{AC}\right) > I_{AC} \quad \forall \quad I_{AC} \in [0, 1). \tag{3.28}$$

The rate of the code in terms of the degree distributions is given as $R = 1 - \frac{1}{d_c \sum_{i=1}^{D_v} \frac{\omega_i}{i}}$. Thus, for given channel conditions, the rate should be maximized such that the error free condition in (3.28) is met. If the check node degree is fixed, maximizing the rate is equivalent to maximizing $\sum_{i=1}^{D_v} \frac{\omega_i}{i}$, which is linear in terms of the design coefficients ω_i. In addition, we have the trivial linear constraints stating that the coefficients

should be non-negative and should to sum to one. Thus the optimization can be solved using linear programming.

b. Design of LDPC code for source transmission

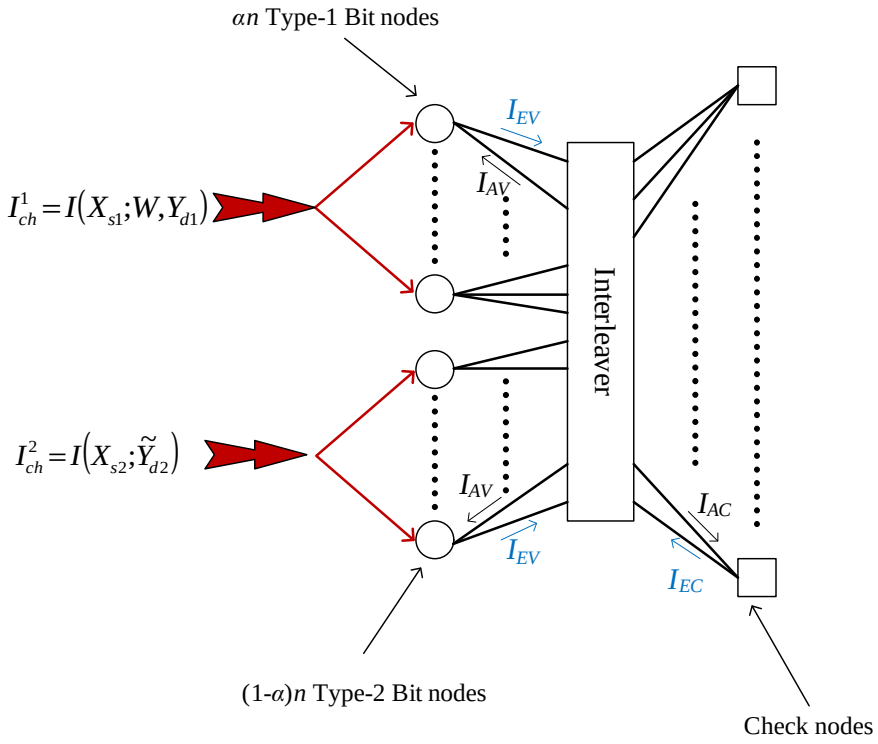


Fig. 24. Information flow in the LDPC code decoding graph for the source transmission.

As mentioned earlier, the LDPC decoding graph for the source transmission has two groups of bit nodes. Typically, the characteristics of the received signal corresponding to the two bit node types are very different from each other, e.g., the SNR for type-1 nodes is much larger than the one for type-2 nodes. Thus, in the design process, we consider different degree distributions for the two bit node types. Let $\Omega_i^{1(2)}$ be the fraction of degree i bit nodes of type 1 (2). Since the fraction of

type-1 bit nodes is α and that of type-2 is $\overline{\alpha}$, we have the following constraints

$$\sum_{i=2}^{D_b} \Omega_i^1 = \alpha; \qquad \sum_{i=2}^{D_b} \Omega_i^2 = \overline{\alpha}, \tag{3.29}$$

where D_b is the maximum bit node degree. Also let $\omega_i^{1(2)}$ be the fraction of total edges connected to degree i variable nodes of type 1(2). Then we have the relationship

$$\begin{aligned} \omega_i^{1(2)} &= \frac{\text{Edges connected to deg } i \text{ nodes of type } 1(2)}{\text{Total edges}} \\ &= \frac{\Omega_i^{1(2)} i}{\sum_{j=2}^{D_b} (\Omega_j^1 + \Omega_j^2) j}. \end{aligned} \tag{3.30}$$

Summing over i and the two types of bit nodes and since $\sum_{i=2}^{D_b} (\Omega_i^1 + \Omega_i^2) = 1$, we get

$$\sum_{i=2}^{D_b} \frac{\omega_i^1 + \omega_i^2}{i} = \frac{1}{\sum_{i=2}^{D_b} (\Omega_i^1 + \Omega_i^2) i}. \tag{3.31}$$

By summing (3.30) over i and using (3.31), we can translate the constraints in (3.29) to a constraint on the degree distributions $\omega_i^{1(2)}$'s as

$$\sum_{i=2}^{D_b} \overline{\alpha} \frac{\omega_i^1}{i} - \alpha \frac{\omega_i^2}{i} = 0. \tag{3.32}$$

In addition, the degree distribution coefficients should obviously sum to one, i.e.,

$$\sum_{i=2}^{D_b} \omega_i^1 + \omega_i^2 = 1. \tag{3.33}$$

The LDPC decoding graph along with the information flows is shown in Fig. 24. As in the previous subsection, let $I_{AC} \in [0, 1]$ be the average a-priori input information to the check nodes. The extrinsic information from the check nodes can be approximated as

$$I_{EC}\left(I_{AC}\right) \approx 1 - \sum_{j=2}^{D_c} \rho_j J\left((j-1) J^{-1}(1 - I_{AC})\right), \tag{3.34}$$

where ρ_j is the fraction of edges connected to check nodes of degree j, and D_c is the maximum check node degree.

As for the EXIT function for the bit nodes, the a-priori information to the bit nodes is the extrinsic information from the check nodes, i.e., $I_{AV} = I_{EC}$. Let $I_{ch}^1 = I(X_{s1}; W, Y_{d1})$ be the input information from the channel for type-1 bit nodes. Similarly, let $I_{ch}^2 = I(X_{s2}; \tilde{Y}_{d2})$ be the input information from the channel for bit nodes of type 2, where $\tilde{Y}_{d2}$ is obtained after hard interference cancellation. Note that we have assumed here that the IRA codeword (comprising of $\mathbf{W}$ and $\mathbf{X}_r$) has been decoded perfectly before the LDPC decoding begins, and hence there is no need to employ soft interference cancellation. The assumption is valid since the design process for the IRA code degree distributions in the previous subsection guarantees (in the Gaussian assumption sense) zero bit error rate for asymptotically large block lengths. In SPA decoding, the extrinsic message on an edge from a bit to check node is the sum of the incoming check to bit node messages on the adjacent edges, and hence, the average extrinsic information from the bit to check nodes is given by

$$I_{EV}\left(I_{AC}\right) = \sum_{k=1}^{2} \sum_{i=2}^{D_v} \omega_i^k J\left(J^{-1}(I_{ch}^k) + (i-1)J^{-1}(I_{EC}(I_{AC}))\right). \tag{3.35}$$

For an error probability zero, we should have [64]

$$I_{EV}\left(I_{AC}\right) > I_{AC} \quad \forall \quad I_{AC} \in [0,1). \tag{3.36}$$

The rate of the code in terms of the degree distributions is given by $R = 1 - \frac{\sum_{j=2}^{D_c} \frac{\rho_j}{j}}{\sum_{i=2}^{D_b} \frac{\omega_i^1 + \omega_i^2}{i}}$. For given channel conditions, the rate should be maximized such that the error free condition in (3.36) is met. If the check node degree distribution is fixed, maximizing the rate is equivalent to maximizing $\sum_{i=2}^{D_b} \frac{\omega_i^1 + \omega_i^2}{i}$, which is linear in terms of the coefficients $\omega_i^{1(2)}$. In addition, the constraints (3.32), (3.33), and (3.36) that need to be satisfied are all linear in terms of these coefficients. Hence this constrained

optimization problem also can be solved using linear programming.

E. Simulation Results

In all our simulations, we assume that the distribution of the binary source messages is uniform. Two topological channel setups, as depicted in Fig. 18, are considered: one with $d_{sr} = 9.5$ and $d_{rd} = 3.15$; another with $d_{sr} = 10$ and $d_{rd} = 2.5$. In each setup, we fix P_r and P_s before finding the information-theoretic optimal half-duplexing parameter α, and the power allocations P_{s1} and P_{s2} required by the simplified SWCNSQ scheme to achieve a given transmission rate R. For both channel setups, Table II lists the optimal parameters for $R = 0.5$ b/s, together with the resulting relay to destination link capacity $C_{rd} = \frac{\overline{\alpha}}{\alpha} C(c_{rd}^2 \tilde{P}_r, c_{sd}^2 P_{s2}, 1)$, the achievable rate $R_1 = C(c_{sd}^2 P_{s1}) + C_r$ on sub-channel 1, and the achievable rate $R_2 = C(c_{sd}^2 P_{s2})$ on sub-channel 2.

Table II. Optimized parameters for simplified SWCNSQ scheme for $R = 0.5$ b/s and relay power $P_r = -12$ dB

d_{sr}	d_{rd}	α	P_s (dB)	P_{s1} (dB)	P_{s2} (dB)	C_{rd}	R_1	R_2
9.5	3.15	0.51	-1.10	-0.047	-2.561	0.783	0.676	0.317
10	2.5	0.54	-1.03	-0.361	-1.976	0.8252	0.628	0.35

Note that if the above information theoretic parameters are used to design LDPC and IRA codes, the rates of the optimized codes will be less than those required because of coding losses. Thus, we fix α and P_{s2} at their theoretical optimum, and keep increasing P_{s1} until the optimization procedure for the CS scheme yields codes of required rates, i.e., we gradually increase P_{s1} until a rate-α IRA code and a rate-$R = \alpha R_1 + \overline{\alpha} R_2$ LDPC code are obtained. We fix P_{s2} at its theoretical minimum and only increase P_{s1} to minimize interference for the IRA decoder as explained in

Section D-1. After good IRA and LDPC codes have been designed, we use DDE [66] to find the minimum transmission power required by the iterative SPA decoder to achieve a close to zero BER when block length and the number of iterations approach infinity. The degree distributions are then used to simulate performance for a finite block length of $n = 2 \times 10^5$ for transmission rates of 0.75, 0.5 and 0.25 b/s. In the following, we summarize our results for the two setups mentioned above.

1. Setup I: $d_{sr} = 9.5$, $d_{rd} = 3.15$, and $P_r = -12$ dB

The optimized degree distributions for the LDPC code at a transmission rates of $R = 0.25, 0.5, 0.75$ b/s are shown in Table III. The BER performance of the degree distributions for $R = 0.5$ b/s is given in Fig. 25. We observe that the CS scheme using IRA and LDPC codes of asymptotic block lengths suffers a coding loss of only 0.17 dB compared to the theoretical SWCNSQ limit. Also, at a BER of 10^{-5}, the practical CF system with finite length codes loses only 0.34 dB from the theoretical bound.

For comparison, we also simulate BER performance of the MS scheme. The LDPC design process for the MS scheme requires designing two separate codes of rates R_1 and R_2 (for example, the two target rates R_1 and R_2 are given in the second row of Table II), and block lengths αn and $\overline{\alpha} n$, respectively. Using DDE, the performance threshold for MS is found to be 0.05 dB worse than that of CS. Recall that MS performs worse than CS because of two reasons; firstly, because of increased interference caused by a higher P_{s2}, and secondly, because of a smaller block length. Since the threshold is calculated for asymptotically large block lengths, this loss of 0.05 dB is due to the increased interference only. The performance of the MS scheme for finite block length of $n = 2 \times 10^5$ is shown in Fig. 25, which is observed to be 0.28 dB worse than CS. At this finite block length, both reasons mentioned above

Table III. Optimized degree distributions at various rates for setup I: $d_{sr} = 9.5$, $d_{rd} = 3.15$, and $P_r = -12$ dB.

R (b/s)			
0.25	LDPC	$\omega^1(x)$	$0.2816x$
		$\omega^2(x)$	$0.0997x + 0.1711x^2 + 0.0619x^4 + 0.1322x^5 + 0.0678x^{14} + 0.0703x^{15} + 0.1153x^{49}$
		$\rho(x)$	$0.4x^3 + 0.6x^4$
	IRA	$\omega(x)$	$0.3085x + 0.2587x^2 + 0.1748x^3 + 0.0219x^6 + 0.1444x^7 + 0.0005x^{25} + 0.0913x^{27}$
		$\rho(x)$	x^2
0.5	LDPC	$\omega^1(x)$	$0.151385x + 0.059467x^{10} + 0.069372x^{14} + 0.349055x^{49}$
		$\omega^2(x)$	$0.047244x + 0.112977x^2 + 0.117668x^4 + 0.022279x^8 + 0.070544x^{12}$
		$\rho(x)$	$0.2x^9 + 0.8x^{10}$
	IRA	$\omega(x)$	$0.133826x + 0.250839x^2 + 0.302605x^6 + 0.31273x^{49}$
		$\rho(x)$	x^4
0.75	LDPC	$\omega^1(x)$	$0.1776x + 0.3747x^{49}$
		$\omega^2(x)$	$0.0095x + 0.1514x^2 + 0.0319x^5 + 0.0061x^6 + 0.0875x^9 + 0.0766x^{10} + 0.0847x^{49}$
		$\rho(x)$	$0.2x^{21} + 0.8x^{22}$
	IRA	$\omega(x)$	$0.1518x + 0.0428x^2 + 0.2491x^3 + 0.156x^5 + 0.1186x^9 + 0.2691x^{19} + 0.0126x^{24}$
		$\rho(x)$	x^5

contribute to this performance loss.

For transmission rates of $R = 0.75$ and $R = 0.25$ b/s, the gap to SWCNSQ theoretical limit with practical LDPC/IRA codes is only 0.14 and 0.2 dB, respectively, for the asymptotic thresholds, and 0.3 and 0.38 dB for finite-length simulations (at BER of 10^{-5}). The three operating points of the practical CS SWCNSQ scheme at a BER of 10^{-5} in comparison with the theoretical limits of DF and direct transmission are shown in Fig. 26(a). As can be seen the operating points at rates 0.75, 0.5, and 0.25 b/s require 0.53, 0.37 and 0.68 dB less power than the DF bound, respectively.

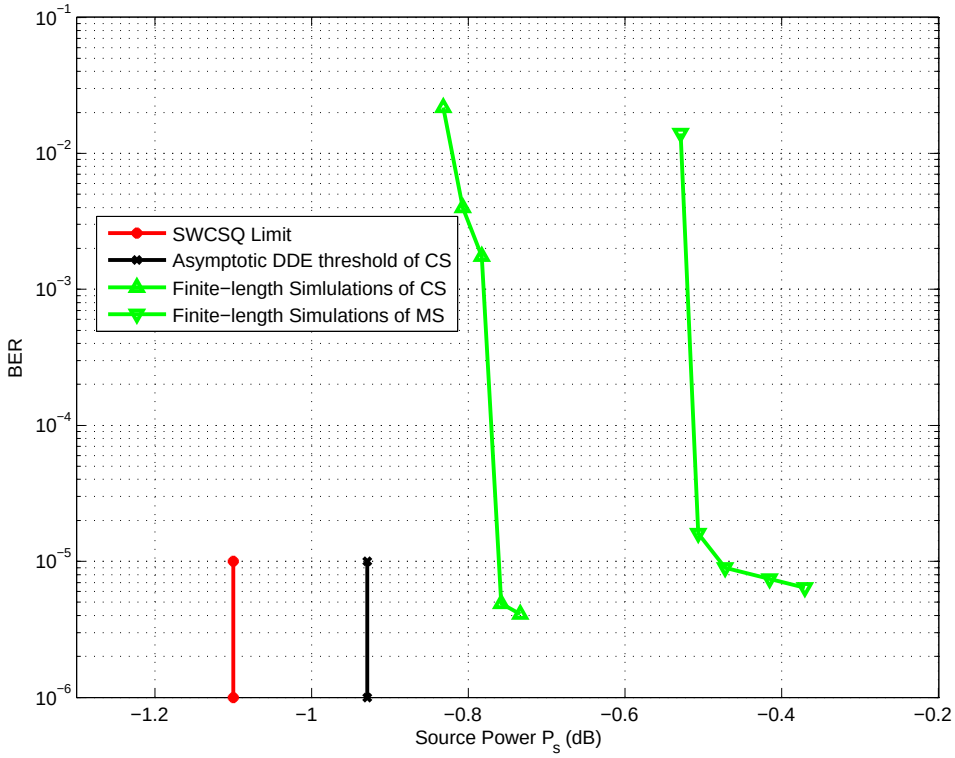
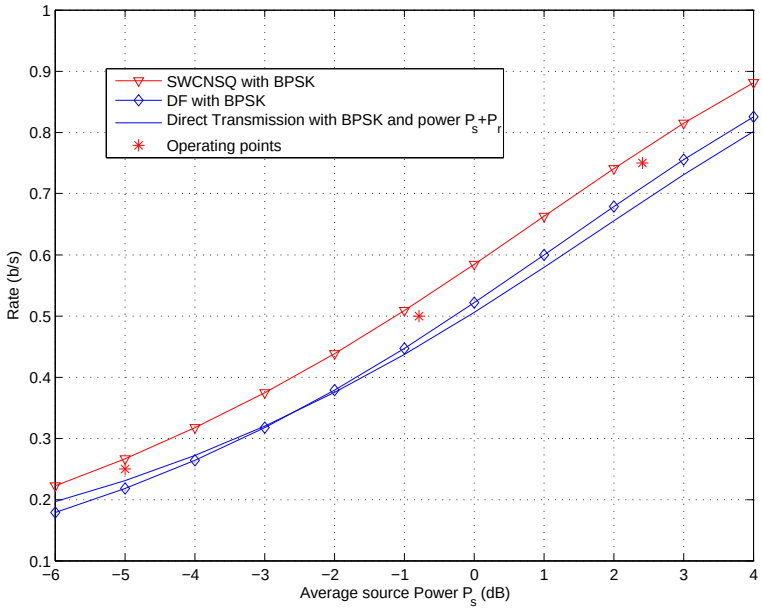


Fig. 25. BER versus the average source power P_s for a transmission rate of $R = 0.5$ b/s, $d_{sr} = 9.5$, $d_{rd} = 3.15$, and $P_r = -12$ dB.

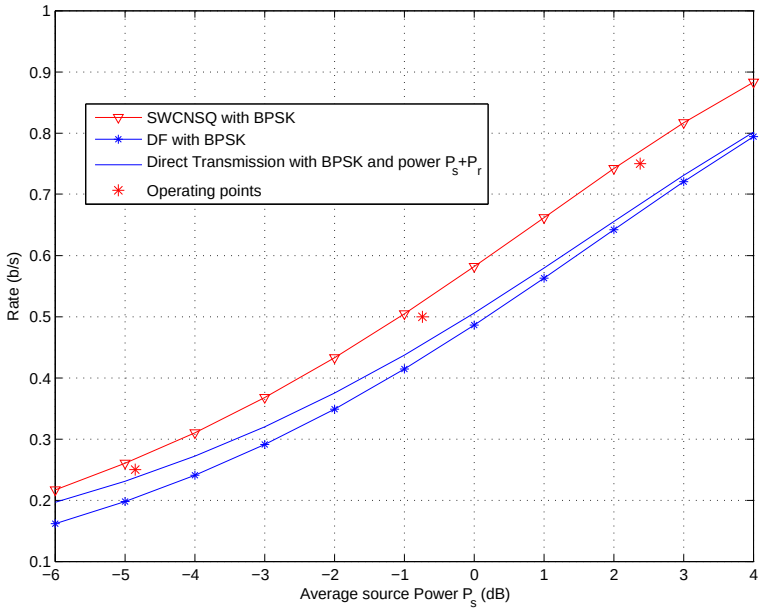
The corresponding gap to the direct transmission limit is 0.86, 0.7, and 0.47 dB, respectively.

2. Setup II: $d_{sr} = 10$, $d_{rd} = 2.5$, and $P_r = -12$ dB

We design three sets of LDPC/IRA codes for transmission rates of 0.75, 0.5 and 0.25 b/s in this new setup − one set of codes for each rate. The optimized degree distributions are shown in Table IV. The asymptotic thresholds are observed to be 0.11, 0.17 and 0.21 dB away from the SWCNSQ limit, respectively. The performance gaps at BER of 10^{-5} with $n = 2 \times 10^5$ are 0.27, 0.33 and 0.38 dB, respectively. Fig. 26(b) shows the operating points at BER of 10^{-5} along with the theoretical bounds. Compared to the the DF limit, the operating points with finite-length simulations



(a)



(b)

Fig. 26. Operating points at BER of 10^{-5} for R=0.25, 0.5 and 0.75 b/s in comparison with DF and direct transmission limits. The relay position corresponds to (a) $d_{sr} = 9.5$ and $d_{rd} = 3.15$, (b) $d_{sr} = 10$ and $d_{rd} = 2.5$. The average relay power is set at $P_r = -12$ dB.

perform better by 1.02, 0.92 and 1.03 dB, respectively, whereas the gap to the direct transmission limit is 0.89, 0.65 and 0.31 dB, respectively.

Table IV. Optimized degree distributions at various rates for setup II: $d_{sr} = 10$, $d_{rd} = 2.5$, and $P_r = -12$ dB.

R (b/s)			
0.25	LDPC	$\omega^1(x)$	$0.2991x$
		$\omega^2(x)$	$0.061x + 0.2393x^2 + 0.0631x^6 + 0.1501x^7 + 0.0187x^{26} + 0.1121x^{27} + 0.0127x^{59} + 0.0438x^{60}$
		$\rho(x)$	$0.4x^3 + 0.6x^4$
	IRA	$\omega(x)$	$0.4797x + 0.198x^2 + 0.0917x^3 + 0.0743x^7 + 0.0799x^9 + 0.0672x^{35} + 0.009x^{36} + 0.0002x^{37}$
		$\rho(x)$	x^2
0.5	LDPC	$\omega^1(x)$	$0.1593x + 0.2052x^{14} + 0.0058x^{15} + 0.3127x^{49}$
		$\omega^2(x)$	$0.0282x + 0.1521x^2 + 0.0352x^5 + 0.1014x^6$
		$\rho(x)$	$0.2x^9 + 0.8x^{10}$
	IRA	$\omega(x)$	$0.3345x + 0.1619x^2 + 0.1883x^3 + 0.0432x^6 + 0.0798x^8 + 0.0628x^{12} + 0.1295x^{25}$
		$\rho(x)$	x^3
0.75	LDPC	$\omega^1(x)$	$0.1893x + 0.0685x^2 + 0.1775x^{49}$
		$\omega^2(x)$	$0.1312x^2 + 0.0184x^3 + 0.0538x^4 + 0.1482x^{10} + 0.0607x^{11} + 0.1524x^{49}$
		$\rho(x)$	$0.2x^{18} + 0.8x^{19}$
	IRA	$\omega(x)$	$0.203x + 0.1424x^2 + 0.2228x^4 + 0.0316x^{11} + 0.1748x^{12} + 0.2255x^{49}$
		$\rho(x)$	x^5

F. Summary

In this chapter, we have considered CF coding with BPSK modulation for the half-duplex Gaussian relay channel. Specifically, we propose SWCNSQ as a means of implementing WZC based CF relaying. We derive the achievable rates of this scheme, specifically with BPSK modulation. Using the information-theoretic bound as a

benchmark, we have developed the first limit-approaching practical CF code design using LDPC codes at the source, and nested scalar quantization and distributed joint source-channel coding with IRA codes at the relay. We have shown that after the quantization indices have been recovered, the destination effectively receives the source transmission over two parallel sub-channels. We use EXIT charts strategy along with a Gaussian approximation to design good degree distributions for LDPC decoding over the two sub-channels. Using density evolution, we find that our code design with an asymptotically large block length is only $0.11-0.2$ dB away from the theoretical limit, whereas finite-length simulation exhibit a gap of only $0.27 - 0.38$ dB from the achievable SWCNSQ bound.

CHAPTER IV

THE QUASI-STATIC FADING RELAY CHANNEL

A. Introduction

In the previous chapter we considered a half-duplex Gaussian relay channel, where all channel coefficients were fixed. In this chapter, we go a step further by considering a half-duplex wireless relay channel where all links experience independent quasi-static Rayleigh flat fading. We assume that the instantaneous fading realizations are not available at the transmitters but are known perfectly at the destination. We focus our attention to the case where the relay network does not have stringent delay requirements. In addition, we assume that it is imperative for the destination to always correctly decode the information being transmitted from the source. In other words, the source and/or the relay can continue to transmit until the destination successfully decodes the message from the source. Since the source does not know beforehand the rate it should use to encode its message, it is impossible to use fixed-rate codes at the source. A similar situation applies at the relay, where it does not know (a) the rate at which to re-encode the message it decoded from the source in case of decode-forward (DF) or (b) the rate at which it should compress and forward what it received from the source in case of compress-forward (CF) relaying. As a solution, we propose rateless coded versions of DF and CF strategies, in which both the source and the relay continue to transmit, using rateless codes, until an acknowledgement (`ACK`) is received from the destination. Since in DF, the relay first decodes the message from the source before encoding it again and transmitting to the destination, the need for rateless coding at the relay is quite obvious. On the other hand, recall from Chapter III that for the case of CF, the signal received at the relay is

source-channel coded before being transmitted to the destination – thus the need for rateless coding might not be so clear. As shown in Chapter III, one can use a single channel code to jointly provide compression (source coding) as well as error protection (channel coding). Thus, in this chapter, we propose using a single rateless channel code to continuously source-channel code the signal at the relay until its decoding at the destination is successful, with the rate compatibility of the underlying rateless channel code lending itself nicely to developing a rate-compatible joint source-channel code. Here we consider it important to mention that whereas rateless coded DF has been explored in the past [75, 76, 77, 78], we believe that this work is the first in the literature which designs and implements a rateless coded CF relaying strategy.

For the fixed channel case, either CF or DF will outperform the other. However, in order to achieve superior performance for the fading case, it is natural that the system should switch to the scheme which results in a higher achievable rate for the current channel conditions. In order to let the system switch between the two relaying schemes, we propose a simple protocol which involves an additional bit of feedback from the destination (in addition to the `ACK` signals) which lets the relay know which scheme to employ. For the case of CF, the time at which this feedback is sent also helps the relay determine the optimum (or near optimum) number of received symbols it should compress (using Wyner-Ziv coding) before starting to transmit to the destination. Assuming the presence of this feedback, we derive achievable rates for rateless coded DF and CF strategies specifically with binary phase-shift keying (BPSK). Next, we implement the rateless coded relaying protocol using Raptor codes [79]. It was shown in [80] that Raptor codes are not universal over additive white Gaussian noise (AGWN) channels, i.e., the optimum degree distribution of the underlying decoding graph is a function of the channel signal-to-noise ratio (SNR). Since the transmitters do not have knowledge of the instantaneous channel SNRs, the source and the relay

do not have the luxury of using degree profiles optimized specifically for the current channel conditions. Instead, we propose a practical approach of using the same degree profiles for all channel conditions, with the degree profiles designed to maximize the throughput averaged over the ensemble of fading realizations. We identify the challenges associated with designing degree distributions for the fading relay channel setting, and formulate the design as a non-linear but convex optimization problem which can be solved using standard iterative convex solvers. Results indicate that with DDE [66] for asymptotically large block lengths, the optimized codes lose only $0.024 - 0.0299$ ($\sim 5\%$) bits from the theoretical limit. On the other hand, the performance loss with finite length codes is observed to be only $0.039 - 0.0528$ ($\sim 9\%$) bits from the theoretical average throughput.

The chapter is organized as follows: We first present the system model in Section B. The relaying protocol which utilizes a combination of rateless coded DF and CF schemes is then described in Section C along with the corresponding performance limits. We then describe the implementation of this rateless coded protocol using Raptor codes and the code design in Section D. The performance with Raptor codes is provided in Section E, and finally a summary of the chapter is provided in Section F.

B. System Model [1]

We consider the same three node as relay model as in Fig. 11, reproduced here as Fig. 27. We assume that all channel coefficients are real − extension of the coding schemes to complex coefficients is straightforward. All channels are assumed to experience quasi-static Rayleigh flat fading, i.e., the channel coefficients change independently

[1]The material presented in this section has some overlap with that in Section B of Chapter III – we will repeat that material here for the reader's convenience.

from one block of transmission to the next with the squares of the absolute values following an exponential distribution whose means are determined by the respective distances. Without loss of generality, we assume that $E[c_{sd}^2] = 1$, $E[c_{sr}^2] = \left(\frac{d_{sd}}{d_{sr}}\right)^l$, and $E[c_{rd}^2] = \left(\frac{d_{sd}}{d_{rd}}\right)^l$, where l is the path loss coefficient. In addition, all links experience additive white Gaussian noises of unit variances. As before, we assume that the path loss coefficient is $l = 3$ and that the source-to-destination distance is fixed at $d_{sd} = 10$.

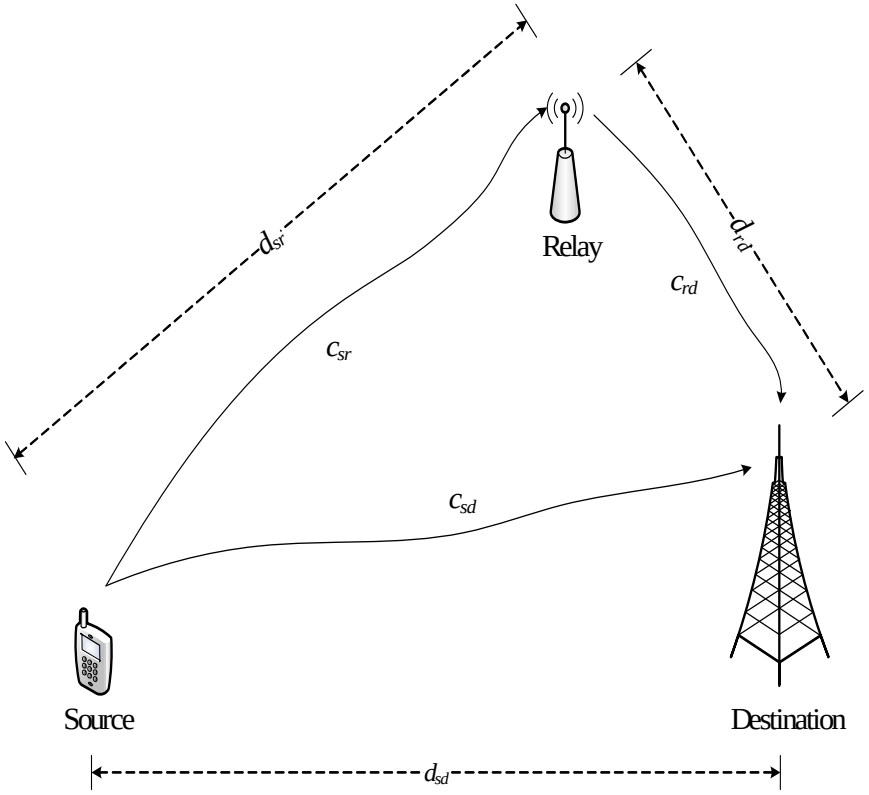


Fig. 27. The relay channel with three nodes: the source, the relay, and the destination.

If the total number of symbols required for successful decoding is N, the relay receives $N_1 = \alpha N$ symbols during the relay receive period T_1, ($\alpha \in [0, 1]$). During T_1,

the source broadcasts its transmission $X_{s1}[n]$, $n = 1, \ldots, \alpha N$, to the relay and the destination which receive

$$Y_r[n] = c_{sr}X_{s1}[n] + Z_r[n], \quad \text{and} \tag{4.1}$$

$$Y_{d1}[n] = c_{sd}X_{s1}[n] + Z_{d1}[n], \tag{4.2}$$

respectively, where $Z_r[n]$ and $Z_{d1}[n]$ are i.i.d. Gaussian noises of unit variances, independent of each other. The relay then transmits for the remaining $\overline{\alpha}N$ symbols during the period T_2 referred to as the relay transmit period, and where $\overline{\alpha} \equiv 1 - \alpha$. During T_2, the destination receives the source transmission $X_{s2}[n]$ and the relay transmission $X_r[n]$ over a multiple-access channel (MAC) with the received signal given by

$$Y_{d2}[n] = c_{sd}X_{s2}[n] + c_{rd}X_r[n] + Z_{d2}[n], \qquad n = 1, \ldots, \overline{\alpha}N, \tag{4.3}$$

where once again $Z_{d2}[n]$ is an i.i.d. additive Gaussian noise of unit variance. Let the source transmission power during T_1 and T_2 be denoted by $P_{s1} = \frac{1}{\alpha N}\sum_{n=1}^{\alpha N} |X_{s1}[n]|^2$ and $P_{s2} = \frac{1}{\overline{\alpha}N}\sum_{n=1}^{\overline{\alpha}N} |X_{s2}[n]|^2$, respectively, and the relay transmission power during T_2 be $P_r = \frac{1}{\overline{\alpha}N}\sum_{n=1}^{\overline{\alpha}N} |X_r[n]|^2$. We consider the following total average system power constraint

$$\alpha P_{s1} + \overline{\alpha}(P_{s2} + P_r) \leq P, \tag{4.4}$$

where P is the maximum allowed average system power per block of transmission. Note that this power constraint is different from the one in Chapter III – we now consider the power budget to be a little more flexible in the sense that one can now also allocate power amongst the source and the relay. If the channel state information (CSI) was available at the transmitters, it would have been possible to optimize the power allocation for every channel realization, i.e., one could have optimized over P_{s1},

P_{s2} and P_r to maximize the overall achievable rate under the total power constraint (4.4). However, since we assume that the CSI is not available at the transmitters, the system cannot reap the benefit from such a power allocation strategy. Instead, we assume that the system employs the same total power in both time slots, i.e., $P_{s1} = P_{s2} + P_r = P$. In addition, the power allocation between the relay and source during T_2 cannot be a function of the channel realizations. Instead, it is a function of the channel statistics only with the specific power allocation chosen to maximize a certain performance criterion such as the average throughput.

C. Ratless Coded Relaying Protocol and Performance Limits

In this section, we will first separately describe the proposed rateless coded DF and CF schemes, along with the corresponding information theoretical performance limits specifically with BPSK modulation. We will then explain the proposed relaying protocol which uses a combination of DF and CF schemes by employing an additional bit of feedback from the destination.

1. Decode-forward

The rateless coded DF scheme we use is similar to the one in [75], except that we have a system power constraint as opposed to separate power constraints for the source and the relay. The source encodes the message sequence $\mathbf{m}$ using a binary rateless codebook of power $P_{s1} = P$. If the source uses a capacity achieving codebook, the relay can decode the message from the source after N_1 symbols if [81]

$$k \leq N_1 C(c_{sr}^2 P), \tag{4.5}$$

where k is the length of the information sequence $\mathbf{m}$, and $C(s)$ is the BPSK channel capacity with SNR s (3.4). It is obvious that the relay can decode the source message before the destination does if and only if (iff) the capacity on the source-to-relay link is higher than that on the source-to-destination link, i.e., $N_1 < N$ iff $|c_{sd}| < |c_{sr}|$. In such a case, after successfully decoding the source message, the relay starts to encode the decoded message using an independent binary rateless codebook of power P_r. Meanwhile, the source continues to ratelessly encode the message using power P_{s2} such that $P_{s2} + P_r = P$. Thus the destination receives $C(c_{sd}^2 P)$ bits of information per symbol during the relay receive period, and $C(c_{sd}^2 P_{s2} + c_{rd}^2 P_r)$ bits per symbol during the relay transmit period. Since we have the power constraint $P_{s2} + P_r = P$, the optimum power allocation strategy during T_2 allocates all power to the node with better channel quality – the other node does not transmit. Thus, if $|c_{sd}| < \min(|c_{sr}|, |c_{rd}|)$, all power is allocated to the relay during T_2, and the destination can decode the message after N symbols if

$$k \leq N_1 C(c_{sd}^2 P) + (N - N_1) C(c_{rd}^2 P). \tag{4.6}$$

On the other hand, if $|c_{sd}| \geq \min(|c_{sr}|, |c_{rd}|)$, either the relay cannot decode the message before the destination, or if it does, it is best for it not to transmit. In either case, relaying does not help at all and the number of symbols N required by the destination for successful decoding satisfies

$$k \leq N C(c_{sd}^2 P). \tag{4.7}$$

The overall achievable rate $R_{DF} = \frac{k}{N}$ as a function of the half-duplexing parameter $\alpha = \frac{N_1}{N}$, can be obtained by combining the constraints (4.5), (4.6), and (4.7) and is

given by

$$R_{DF} = \begin{cases} C(c_{sd}^2 P) & \text{if } |c_{sd}| \geq \min\left(|c_{sr}|, |c_{rd}|\right), \\ \max_\alpha \left[\min\left(\alpha C(c_{sd}^2 P) + \overline{\alpha} C(c_{rd}^2 P), \alpha C(c_{sr}^2 P) \right) \right], & \text{otherwise.} \end{cases} \tag{4.8}$$

The achievable rate expressions in (4.8) can further be simplified as follows. Since $|c_{sd}| < |c_{rd}|$ in the second line of (4.8), the first argument of the min function decreases while the second argument increases with α. Thus the optimum choice of the half-duplexing parameter $\alpha^\dagger = \frac{C(c_{rd}^2 P)}{C(c_{rd}^2 P) + C(c_{sr}^2 P) - C(c_{sd}^2 P)}$ in (4.8) is the solution obtained by equating the two arguments. When $|c_{sr}| \leq |c_{sd}|$, i.e., when DF relaying reduces to the direct transmission case, $\alpha^\dagger \geq 1$. Thus if $\alpha_{DF}^* = \min(\alpha^\dagger, 1)$, we have

$$R_{DF} = \alpha_{DF}^* C(c_{sd}^2 P) + (1 - \alpha_{DF}^*) C(\max(c_{rd}^2, c_{sd}^2) P). \tag{4.9}$$

The overall rateless coded DF relaying can be summarized as follows. The source starts to encode its message with a ratless codebook using power P. If $|c_{sd}| \geq \min\left(|c_{rd}|, |c_{sr}|\right)$, the relay does not attempt to decode and the source continues its transmission until an `ACK` has been received from the destination. Otherwise, the relay attempts to decode and sends an `ACK` to the source once it is successful, at which point the source stops transmitting. The relay now starts encoding using a rateless codebook with power P and continues transmitting until an `ACK` is received from the destination.

2. Compress-forward

For the CF scheme, the source initially encodes the message using a binary rateless code with power $P_{s1} = P$. After N_1 symbols have been received at the relay, it quantizes $\mathbf{Y}_r$ to $\mathbf{W}$ and then employs Slepian-Wolf (SW) coding [15] to compress the random sequence $\mathbf{W}$ with the sequence $\mathbf{Y}_{d1}$ treated as the side-information known

at the decoder, but not at the encoder. During the relay transmit period T_2, the compressed version is coded using a rateless channel code and transmitted to the destination using power P_r. Meanwhile, the source continues to transmit using a rateless codebook of power P_{s2} with $P_{s2} + P_r = P$. Let the achievable rates during T_2, lying inside the MAC capacity region [17], on the relay-to-destination and source-to-destination link be R_{rd} and R_{sd}, respectively. Then the destination should be able to recover $\mathbf{W}$ after a total of N_R symbols have been transmitted if [71]

$$N_1 H(W|Y_{d1}) \leq (N_R - N_1) R_{rd}, \tag{4.10}$$

where the italicized letter represents the random variable associated with the i.i.d. sequence denoted by the same boldface letter. After $\mathbf{W}$ has been successfully recovered, it is used to extract information about $\mathbf{m}$. Thus the destination receives $I(X_{s1}; W, Y_{d1})$ bits of information during T_1, and R_{sd} bits of information during T_2 – the message can be recovered after N symbols if

$$k \leq N_1 I(X_{s1}; W, Y_{d1}) + (N - N_1) R_{sd} \tag{4.11}$$

provided that N_1 and N_R satisfy (4.10) and that $N \geq N_R$ (the destination can recover $\mathbf{m}$ only after $\mathbf{W}$ has been decoded). Dividing both sides of (4.11) by N, using $R_{CF} = \frac{k}{N}$, $\alpha = \frac{N_1}{N}$, expanding the term $I(X_{s1}; W, Y_{d1})$, and maximizing over α, we get the rate for the rateless coded CF relaying scheme as

$$R_{CF} = \max_{0 \leq \alpha \leq 1} \alpha \left(C(c_{sd}^2 P) + H(W|Y_{d1}) - H(W|X_{s1}) \right) + \overline{\alpha} R_{sd}, \tag{4.12}$$

subject to the condition (obtained by diving both sides of (4.10) by N)

$$\alpha H(W|Y_{d1}) \leq \left(\frac{N_R}{N} - \alpha \right) R_{rd}. \tag{4.13}$$

We restrict our attention to the case where $\mathbf{W}$ is the output of a binary quantizer with $\mathbf{Y}_r$ as the input. The primary motivation, as mentioned in Chapter III, is the fact that for BPSK modulated channels, restricting the quantization to be binary suffers an insignificant performance loss. At the same time, it offers great practical advantages for the fading case since the quantizer need not change for varying channel conditions. Thus for our practical scheme, we use the same quantizer as that of the simplified scheme of Section C-5 in Chapter III. Upon observing (4.13), we find that in order to aid the recovery of $\mathbf{W}$ for a fixed quantizer (and thus fixed $H(W|Y_{d1})$) and a fixed N, one must either increase N_R, or decrease α by decreasing N_1. The former can be varied by letting the relay continue to transmit until the destination generates an `ACK`, and the latter with the help of an additional feedback as mentioned in Section C-3. The following result shows the optimum balance between N_R, and α.

Proposition 1 For a fixed quantizer, achieving the best rate in (4.12) requires that the destination recover $\mathbf{W}$ and $\mathbf{m}$ at the same time. In addition, the half-duplexing parameter which maximizes (4.12) is given as

$$\alpha_{CF}^{*} = \frac{R_{rd}}{R_{rd} + H\left(W|Y_{d1}\right)}. \tag{4.14}$$

Proof The first part of the proposition follows from the constraint $N_R \leq N$, implying that the value of N_R for which the the constraint (4.13) is the most relaxed, and hence R_{CF} is maximized is $N_R = N$. The proof of the second part of the proposition is as follows. We have that

$$C\left(c_{sd}^2 P\right) > C\left(c_{sd}^2 P_{s2}\right) \geq R_{sd},$$

where the first inequality is because $P > P_{s2}$, and the second follows from the constraints on the MAC capacity region [17]. In addition, we have

$$H\left(W|Y_{d1}\right) \geq H\left(W|Y_{d1}, X_{s1}\right) = H\left(W|X_{s1}\right),$$

where the inequality appears because conditioning reduces entropy, whereas the equality is because of the fact that given X_{s1}, W is independent of Y_{d1}. We thus have

$$C\left(c_{sd}^2 P\right) + H\left(W|Y_{d1}\right) - H\left(W|X_{s1}\right) > R_{sd},$$

implying that R_{CF} is increasing in α. As for the constraint (4.13), note that the left hand side is increasing in α, while the right hand side is decreasing. Thus, the maximum, and the optimum value of α which maximizes (4.12) while at the same time satisfying (4.13) is obtained by equating the two sides of the constraint, thus giving rise to the value in (4.14).

A loose end that still remains is to determine where R_{sd} and R_{rd} should lie in the MAC capacity region. The optimum operating point in the MAC capacity region is as follows.

Proposition 2 The achievable rate of the CF strategy described above is maximized when R_{sd} and R_{rd} lie on the corner point of the sum-rate side of the MAC capacity region, with the corner point characterized by $R_{sd} = I\left(X_{s2}; Y_{d2}|X_r\right)$ and $R_{rd} = I\left(X_r; Y_{d2}\right)$.

Proof *See Appendix D.*

The above proposition implies that the destination should first attempt to decode the quantized sequence $\mathbf{W}$ using $\mathbf{Y}_{d1}$ as the side-information and by treating the transmission $\mathbf{X}_{s2}$ from the source as binary interference − then the destination should cancel the interference $\mathbf{X}_r$ from $\mathbf{Y}_{d2}$ before attempting to decode the original message $\mathbf{m}$. The result fares well for a practical scheme since one can use separate decoders for $\mathbf{W}$ and $\mathbf{m}$ instead of a more complex joint decoder. Finally, since no CSI is available at the transmitters, the power allocations P_{s2}, P_r cannot change from one channel

realization to the other. Therefore, for some channel realizations, the expression in (4.12) might result in an achievable rate less than the direct transmission strategy, as also indicated in the proof of Proposition 2. However, one can always reduce the CF scheme to a direct transmission strategy by setting $\alpha = 1$ and not attempting to recover $\mathbf{W}$ at the destination. In short, the achievable rate for the rateless coded CF relaying is given by

$$R_{CF} = \max\left\{\alpha^*_{CF}\left(C(c^2_{sd}P) + H(W|Y_{d1}) - H(W|X_{s1})\right) + \overline{\alpha^*_{CF}}C(c^2_{sd}P_{s2}) \ , \ C(c^2_{sd}P)\right\}. \tag{4.15}$$

Note that the above analysis requires the relay to know the optimum number of symbols it should receive and quantize before it starts transmitting. With no CSI, the relay is unable to determine this on its own. However, in the following, we explain how an additional bit of feedback from the destination can convey this information.

3. Overall relaying protocol

The proposed relaying protocol, which uses a combination of rateless coded DF and CF schemes, depending on which results in a higher achievable rate is depicted in Fig. 28. Without full CSI, the source and the relay cannot determine

1. whether to employ DF or CF coding, and

2. the optimum half-duplexing parameter α^* for CF coding (for DF coding without CSI, the relay is still capable of determining when to stop receiving and start transmitting − it does so when it correctly decodes the source message).

On the other hand, since the destination knows all channel coefficients, it is able to determine both of the above. Thus, we propose for the destination to broadcast one additional bit of feedback, referred to as `SCH`, after it has received N_1 symbols –

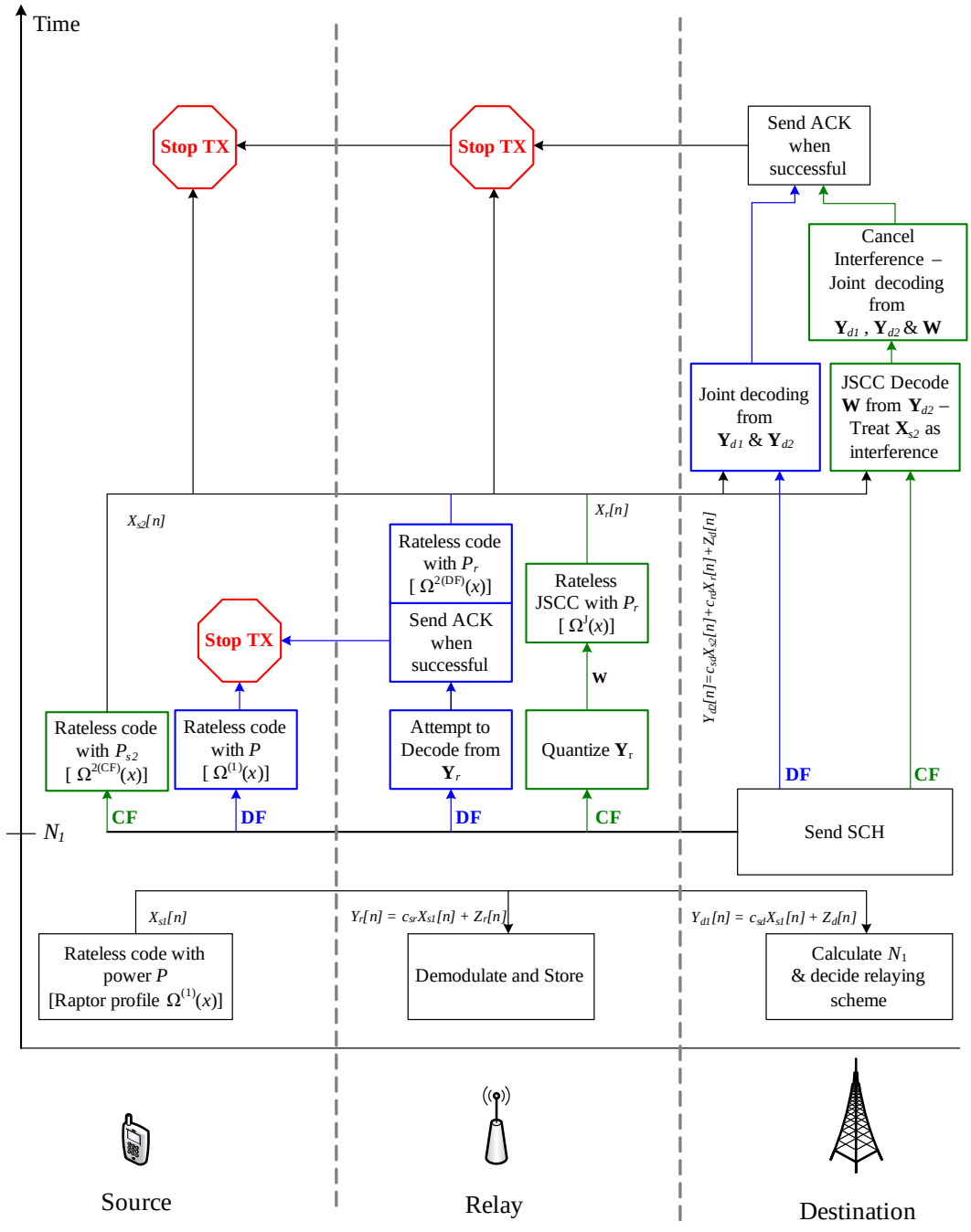


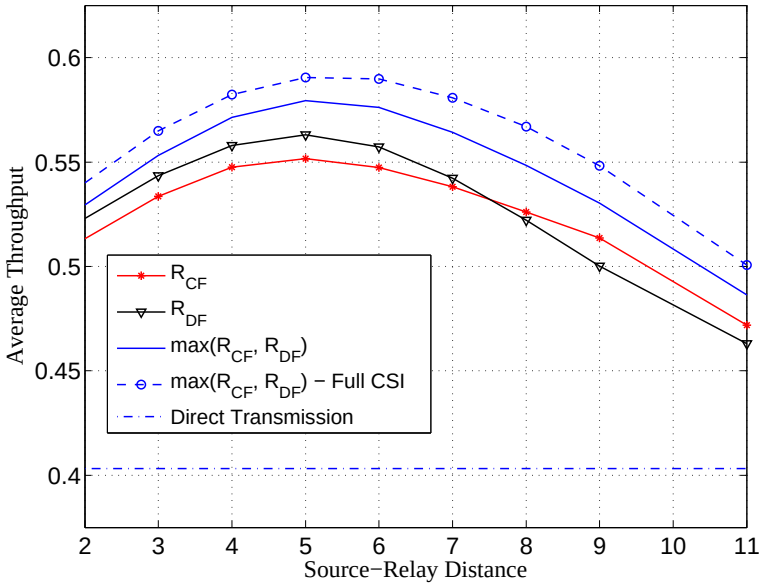
Fig. 28. The overall relaying protocol with no CSI at the transmitters. The destination knows all channel coefficients.

the optimum number of symbols the relay should receive. The signal SCH not only indicates to the source and the relay whether to employ the DF or the CF scheme, but the time at which it is broadcasted also conveys information about the number of symbols the relay should receive before it starts transmitting. The achievable rate of the overall relaying strategy for fixed channel realizations is then given by $R = \max\{R_{DF}, R_{CF}\}$, where R_{DF} and R_{CF} are given by (4.9) and (4.15), respectively.

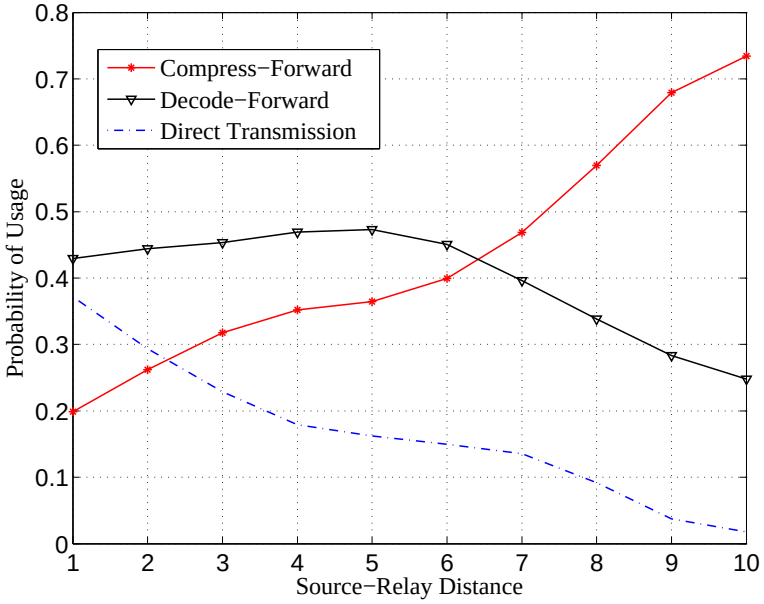
Finally, as mentioned before, for some channel realizations, the best thing to do is not to relay at all. The protocol described above can easily incorporate such a transmission strategy. If the destination determines direct transmission to be the best choice, it never broadcasts SCH. The relay continues to wait for SCH which never comes, and thus the relay never transmits. The destination therefore attempts to decode directly from the source transmissions.

4. Numerical results

As a performance measure, we consider the mean throughput averaged over the fading distributions. In Fig. 29(a), we plot the average throughput as a function of the source-relay distance when $d_{sd} = d_{sr} + d_{rd}$, i.e., the relay is moving along a straight line between the source and the destination, and when the power allocations P_{s2} and P_r are chosen to maximize the average throughput. We point out that we consider the straight line setup for illustration purposes only; the analysis and the coding schemes presented in earlier sections are applicable to any geometrical setting of the nodes. For comparison, we also plot the average throughput for the case where all nodes have perfect knowledge of the instantaneous realizations of all channel coefficients. The achievable rates for the perfect knowledge case can be found by maximizing (4.9) and (4.15) over the power allocations P_{s1}, P_{s2} and P_r such that the power constraint (4.4) is satisfied. As seen from Fig. 29(a), even with no CSI at the transmitters, the relaying



(a)



(b)

Fig. 29. Source-to-relay distance versus (a) average system throughput and (b) probability of usage of each relaying scheme with no channel knowledge at the transmitters. The relay is assumed to be moving along a straight line between the source and the destination and the average system power is set at $P = 0$ dB.

strategy described in Section C-3 achieves a significantly higher average throughput than that with direct transmission. For instance, when $d_{sr} = 6$, relaying outperforms direct transmission by 0.173 bits/sample (b/s). At the same distance, having perfect channel knowledge at the nodes results in an additional gain of 0.013 b/s, where the additional gain is because of the flexibility the perfect channel knowledge allows in the power allocation, as mentioned in Section B. We also provide the average throughput if the system were to employ only a single relaying strategy instead of a combination. It is obvious that using a combination is better than CF or DF employed individually. For example, at $d_{sr} = 6$, the proposed protocol outperforms DF by 0.019 b/s, whereas at $d_{sr} = 8$, it is 0.0224 b/s better than CF.

In Fig. 29(b), we plot the probability of usage of each relaying scheme as a function of d_{sd}, where once again the relay is assumed to moving along a straight line between the source and the destination. When the relay is close to the source, DF is employed more often than CF. However, as the relay moves closer to the destination, the probability of usage of CF relaying increases while that of DF decreases. For example, at a distance $d_{sr} = 9$, the CF scheme is the best choice with probability 0.68, while DF outperforms all others with a lower probability of 0.28. This makes intuitive sense since as the relay moves closer the destination, it is more likely that the source-to-relay channel is weaker, due to which it takes longer, if at all, for the relay to decode the source transmission. On the other hand, with the relay being closer to the destination, it is quite likely that the signals received at the relay and the destination are highly correlated, thus allowing SW coding at the relay to reap the benefits of this correlation. Also note that the probability of usage of the direct transmission strategy is non-zero. Had perfect CSI been available at all nodes, the nodes would have been able to utilize the optimum power allocation due to which the relaying protocol would have always been able to outperform direct transmission.

D. Practical Rateless Coded Relaying with Raptor codes

In this section, we explain how the rateless coded relaying protocol described in Section C can be put into practice using Raptor codes [79]. As mentioned before, we propose using the same degree profiles for all channel conditions, with the degree profiles designed to maximize the average system throughput. The source first precodes its message with a rate-R_p LDPC code. The precoded bits are then encoded with an LT code [82] characterized by a degree distribution polynomial $\mathbf{\Omega}^{(1)}(x) = \sum_{d=1}^{D} \Omega_d^{(1)} x^d$, where D is the maximum node degree. The degree of every output node is chosen randomly, with the probability of choosing a degree d being Ω_d. For a degree d output node, d precoded bits are chosen randomly, which are added modulo-two to obtain the corresponding output bit. The output bits are then BPSK modulated to the constellation points $\left\{+\sqrt{P}, \sqrt{-P}\right\}$ and transmitted to the source as well as to the relay. The source continues to transmit until the relay receive period is over. Since the source does not know beforehand which relaying scheme, if any, will be employed in the future, we let the source during T_1 encode its message using a single profile $\mathbf{\Omega}^1(x)$ regardless of whether DF, CF or direct transmission will be employed. After N_1 symbols have been transmitted, the relay network either employs a DF or a CF scheme, indicated to it by the `SCH` feedback from the destination. In the following, we will first describe separately the implementation of rateless coded DF and CF schemes with Raptor codes – the implementation of the two schemes along with the overall relaying protocol is summarized in Fig. 28. We will then explain the design process for the Raptor code degree distributions which we formulate as a non-linear convex optimization problem.

1. Decode-forward

The implementation of the DF scheme with Raptor codes is fairly straightforward. The destination broadcasts `SCH` after $N_1 = \lceil \alpha^* N \rceil = \lceil \frac{k}{C(c_{sr}^2 P_{s1})} \rceil$ symbols − the theoretical limit for successful decoding. The source continues to transmit until decoding at the relay is successful[2], indicated to the source by an `ACK`. At this point in time, the source stops transmitting and all the power is allocated to the relay. The relay first precodes the decoded information sequence with the same rate-R_p LDPC code that the source was using, and then encodes the precoded bit sequence with an LT code characterized by the degree distribution polynomial $\mathbf{\Omega}^{(2)DF}(x)$. Since both the source and the relay use the same LDPC code to precode the same information sequence, the destination can recover the source message using a joint decoding graph as shown in Fig. 30. As indicated in the figure, the decoding graph of a Raptor/LT code contains two types of bit nodes: (a) the left bit nodes corresponding to the precoded information bits and (b) the right bit nodes corresponding to the LT coded bits transmitted over the channel. Whereas the left bit nodes do not have any a-priori log-likelihood ratios (LLRs), the a-priori LLRs for the right bit nodes are evaluated from the corresponding channel values. If N_1 and $N_2 = N - N_1$ are the number of symbols transmitted from the source and the relay, respectively, the right bit nodes can further be divided into two categories: (a) N_1 Type-1 bit nodes corresponding to the coded bits transmitted from the source during T_1 using degree profile $\mathbf{\Omega}^{(1)}(x)$ and whose channel LLRs are calculated from $\mathbf{Y}_{d1}$ and (b) N_2 Type-2 bit nodes corresponding to the coded bits transmitted from the relay during T_2 using degree profile $\mathbf{\Omega}^{(2)DF}(x)$ and whose channel LLRs are calculated from $\mathbf{Y}_{d2}$. With these channel

[2]We assume that the nodes always know if they have decoded the correct codeword. In practice this can be ensured by using the likes of cyclic redundancy-check codes.

LLRs, the decoder runs iterative belief-propagation (BP) algorithm on the joint decoding graph of Fig. 30 and obtains the extrinsic LLRs on the information bits. For a given number of received symbols N, the BP algorithm runs until either a maximum number of iterations are reached or the correct codeword is decoded. In the case of the former, the decoder waits for more transmissions from the relay before restarting the decoding process. On the other hand, if the correct codeword is decoded, the destination generates an `ACK` and a practical achievable throughput of $\frac{k}{N}$ is recorded.

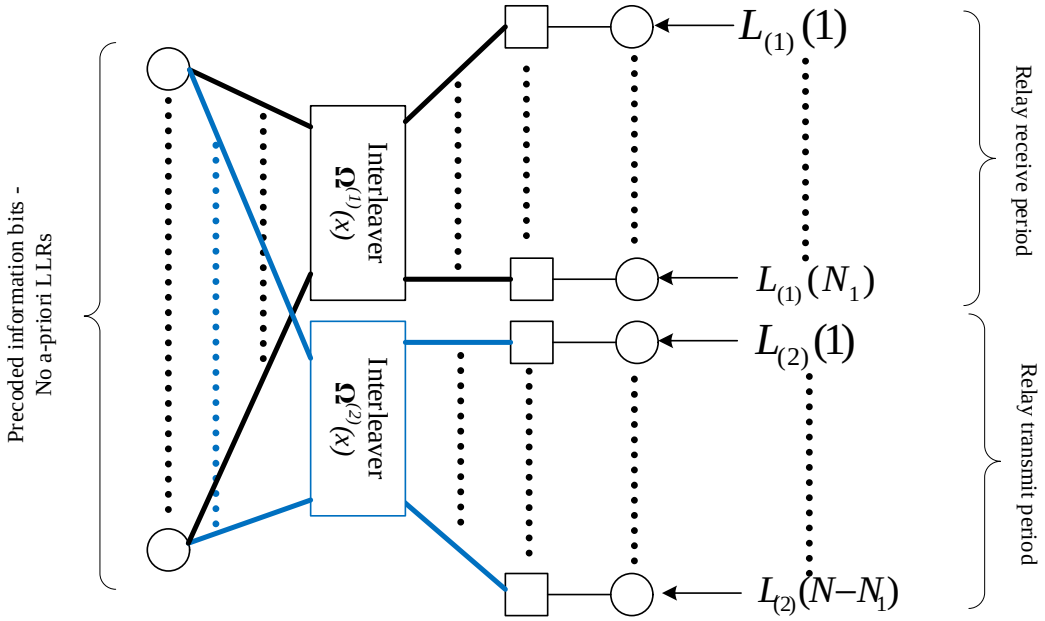


Fig. 30. Joint decoding graph for recovering the original message from the source. The decoding graph for the LDPC precode is not shown here for clarity.

2. Compress-forward

If SCH, transmitted after N_1 symbols, indicates CF to be the best choice, the relay quantizes $\mathbf{Y}_r$ to obtain $\mathbf{W}$. The quantized sequence now needs to be source coded (SW compressed) with $\mathbf{Y}_{d1}$ as the side-information, as well as channel coded against the noise on the relay-to-destination link. As opposed to the fixed-rate distributed joint source-channel coding (DJSCC) for the Gaussian relay channel mentioned in Chapter III, we propose rateless DJSCC using Raptor codes for the fading relay channel. After describing the rateless DJSCC at the relay, we will briefly describe the decoding of the joint source-channel code at the destination, in addition to the decoding of the original message from the source.

a. Rateless DJSCC

The basic idea of rateless DJSCC of the sequence $\mathbf{W}$ using Raptor codes is shown in Fig. 31 (a). The binary sequence $\mathbf{W}$ is first precoded with a systematic LDPC code. The precoded output is encoded with an LT code characterized by the degree distribution polynomial $\mathbf{\Omega}^J(x)$, where we have used the superscript J to indicate joint source-channel coding. The output bits of the LT code are BPSK modulated with power P_r and transmitted over the noisy relay-to-destination channel. As in regular point-to-point Raptor encoding/decoding, the precoded information bits themselves are not transmitted over the physical channel. However, note that the systematic precoded information bits $\mathbf{W}$ are correlated with $\mathbf{Y}_{d1}$ which was received by the destination during T_1 and is therefore already available as the side-information. Thus, one can think of $\mathbf{W}$ as being transmitted over a virtual correlation channel with $\mathbf{Y}_{d1}$ as the output. The decoding graph at the destination for recovering $\mathbf{W}$ is shown in Fig. 31 (b). Treating $\mathbf{X}_{s2}$ as binary interference, the destination uses the received

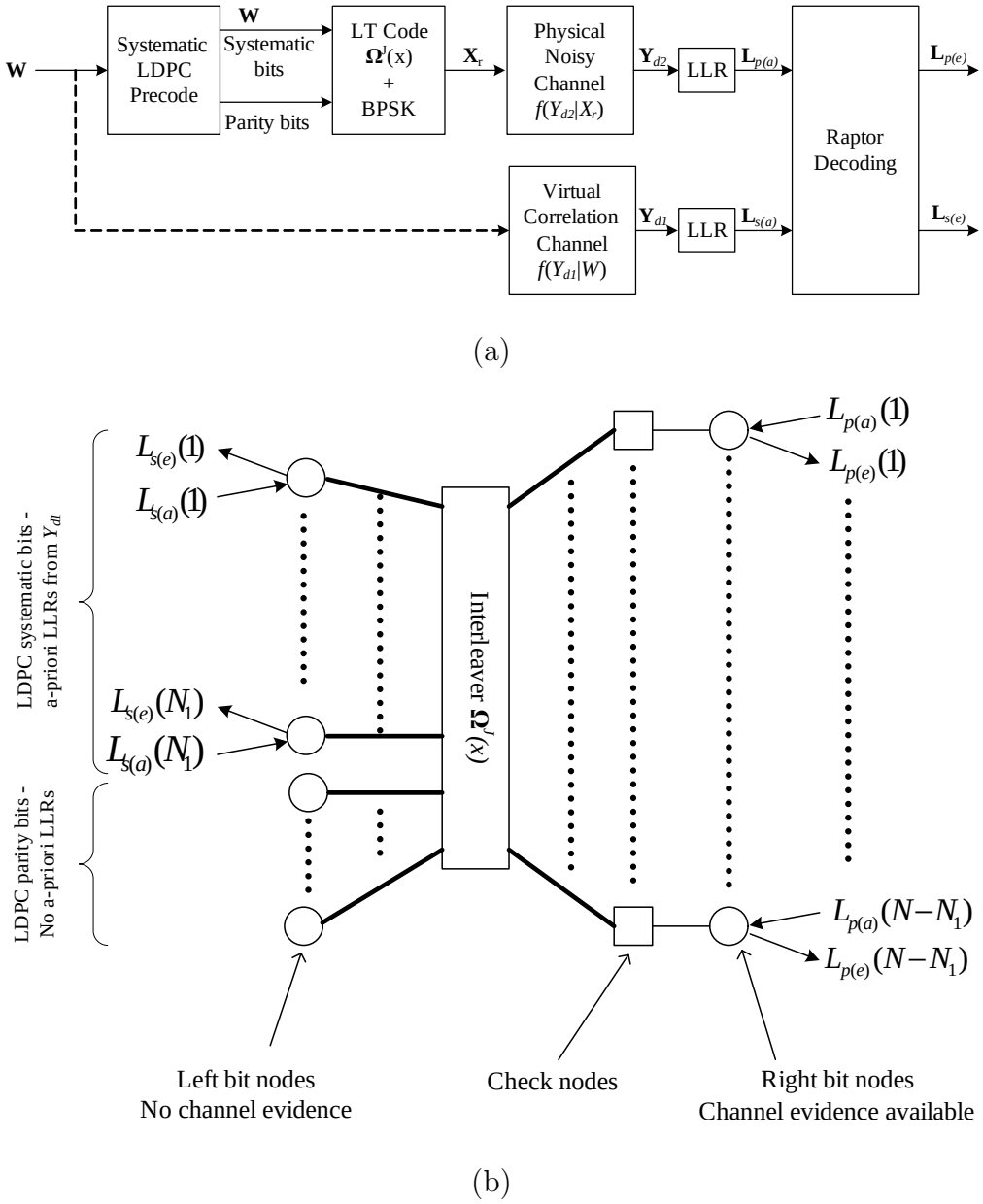


Fig. 31. Rateless DJSCC of $\mathbf{W}$ and the side-information $\mathbf{Y}_{d1}$ using a Raptor code. (a) The LT coded bits are transmitted over the physical noisy channel, while $\mathbf{W}$ is assumed to be transmitted over the virtual correlation channel. (b) Decoding graph for recovering $\mathbf{W}$. $\mathbf{Y}_{d1}$ and $\mathbf{Y}_{d2}$ are used to evaluate the LLRs for the left and right bit nodes, respectively. The decoding graph outputs extrinsic LLRs for both $\mathbf{W}$ and $\mathbf{X}_r$. The decoding graph for the LDPC precode is once again not shown for clarity.

sequence $\mathbf{Y}_{d2}$ to evaluate the a-priori LLRs for the right bit nodes corresponding to $\mathbf{X}_r$. In traditional Raptor decoding, only the right bit nodes have the a-priori LLRs available from the channel − the a-priori LLR inputs to the left bit nodes are zero. On the other hand, for DJSCC, the received sequence $\mathbf{Y}_{d1}$ is correlated with the systematic bits of the LDPC precode, and hence can be used to provide non-zero a-priori LLR values[3]. However, since the LDPC parity bits are, in general, not correlated with $\mathbf{Y_{d1}}$, the a-priori LLRs to the left bit nodes corresponding to these bits will still be zero. After running a predetermined number of BP iterations on the graph, the DJSCC decoder outputs the extrinsic LLRS $\mathbf{L}_{s(e)}$ and $\mathbf{L}_{p(e)}$ for the systematic bits $\mathbf{W}$ and the coded symbols $\mathbf{X}_r$, respectively. These extrinsic LLRs are then used to recover the original information from the source, as will be explained next.

b. Decoding the source information

During the relay transmit period, the source uses power P_{s2} and continues to transmit by LT encoding the precoded information sequence using a degree profile $\mathbf{\Omega}^{(2)CF}(x)$. Thus as in the DF scheme, the right bit nodes of the decoding graph for CF can also be divided into two categories: (a) N_1 Type-1 bit nodes corresponding to T_1 encoded with profile $\mathbf{\Omega}^{(1)}(x)$. Both $\mathbf{Y}_{d1}$ and the decoded version of $\mathbf{W}$ are used to calculate the a-priori LLRs for these bit nodes, and (b) N_2 Type-2 bit nodes corresponding to T_2 encoded with profile $\mathbf{\Omega}^{(2)CF}(x)$. The LLRs for these nodes are calculated from $\mathbf{Y}_{d2}$ after canceling the interference caused by $\mathbf{X}_r$. The LLRs for type-1 bit nodes can be evaluated from $\mathbf{Y}_{d1}$ and the extrinsic information on the systematic nodes of

[3]See (3.18) and (3.20) for exact expressions for the a-priori LLRs to the left and right bit nodes, respectively

the DJSCC decoder [4]. Using these a-priori LLRs the destination runs BP algorithm on the decoding graph for a fixed number of iterations. If the correct codeword is decoded, the destination generates an `ACK` indicating to the source as well as to the relay to stop transmitting. Otherwise, the destination waits for the source and the relay to transmit more symbols and keeps running iterative BP algorithm on the two decoding graphs (for $\mathbf{W}$ and $\mathbf{m}$) until $\mathbf{m}$ is decoded correctly.

Finally, we mention a couple of important points about the rateless coded CF scheme. First, note that the length N_1 of the relay receive period is always fixed for given channel realizations. This is in contrast to the DF scheme, where for a given set of channel realizations, N_1 is variable, at least for finite length codes. This is because for DF, the relay receive period does not end until the relay is able to correctly decode $\mathbf{m}$. Second, recall that Proposition 1 dictates that N_1 should be chosen in such a manner so that the destination recovers the quantized sequence $\mathbf{W}$ and the source message $\mathbf{m}$ at exactly the same time. Another way, perhaps from a practical standpoint, to explain the logic behind this requirement is the following. If N_1 were too small, $\mathbf{W}$ would be recovered before $\mathbf{m}$, and thus the relay would be sending redundant information during the time $\mathbf{W}$ has been recovered but not $\mathbf{m}$. In such a case, one could have clearly picked a larger N_1 such that the new $\mathbf{W}$ would have still been recoverable before $\mathbf{m}$ while at the same time conveying more information about the source transmissions − thus effectively reducing the time it takes to decode $\mathbf{m}$. On the other hand, if N_1 were too large, the time it takes for the destination to decode $\mathbf{W}$ would be large, and might serve as a bottleneck in declaring successful decoding. This is because the destination has to have (almost) successfully decoded $\mathbf{W}$ before attempting to decode $\mathbf{m}$. Clearly, one could have chosen a smaller N_1 in order to

[4]See (3.23) and (3.21) for exact expressions for the a-priori channel LLRs to type-1 and type-2 bit nodes, respectively

reduce the time it takes to decode $\mathbf{W}$ and thus avoid the bottleneck. The optimum balance between N_1 being too small or too large is struck when it is chosen such that $\mathbf{W}$ and $\mathbf{m}$ are recovered at the same time, as indicated by Proposition 1. Besides the information-theoretic analysis, such an optimum choice of N_1 is also possible for asymptotically large code lengths, as will be explained in Section E. However, it is not always possible to choose such an N_1 for finite-length codes, in which case we set it to the optimum value obtained for infinite code lengths.

3. Degree profile design

The design requirement is to choose the degree profiles $\mathbf{\Omega}^{(1)}(x)$, $\mathbf{\Omega}^{(2)DF}$, $\mathbf{\Omega}^{(2)CF}$, and $\mathbf{\Omega}^{J}(x)$ such that the system throughput, averaged over the channel ensemble, is maximized. The design process for optimizing the degree profiles for the fading relay channel is obviously more challenging than that of traditional point-to-point communication. This is because the design methodology must:

1. obtain degree profiles which yield the best *average* performance when used over a variety of channel conditions,

2. cater for different channel conditions on the two types of right bit nodes of the decoding graph (for example, in DF, the SNR for nodes corresponding to T_1 is determined by c_{sd}, while for those correpsonding to T_2, it is determined by c_{rd}), and

3. take into account the non-zero information on the systematic precoded bits in DJSCC decoding – for typical Ratpor code decoding, this information is always zero.

For the design process, we use the EXIT function strategy [64] along with the Gaussian assumption [65], i.e., we assume that all LLRs in the iterative BP decoding have a

symmetric Gaussian distribution. In addition, we use the following approximation

$$J(\mu) \approx 1 - e^{-a\mu}, \tag{4.16}$$

where $a > 0$ is a constant, and $J(\mu)$ is the information conveyed by a Gaussian LLR with mean μ and variance 2μ [64]. The use of this approximation is justified for two reasons. First, the approximation can be shown to be close to the actual function as shown in Fig. 32. Second, we use the approximation only for designing the degree profiles − the actual performance of the designed profiles is evaluated using exact DDE [66] as well as finite length simulations, both indicating operation close to the theoretical limits. In the following, we first describe a condition necessary for successful BP decoding (in the EXIT function and Gaussian assumption sense) for a very general Raptor decoding setup. We will then use this condition to explain the design procedure for the rateless coded relaying schemes by showing that all decoders for the relaying schemes can be considered as specific cases of this general decoder.

a. General decoding setup

For the general setup, we consider the case where there are two types of right bit nodes. For example, one type of right bit nodes can correspond to the relay receive period, and the other type to the relay transmit period. We assume that the check nodes corresponding to these two types of right bit nodes follow different degree distributions (from the node perspective) $\mathbf{\Omega}^{(i)}(x) = \sum_{d=1}^{D} \Omega_d^{(i)} x^d$, $i = 1, 2$. In addition, we assume that both the right as well as the left bit nodes have an a-priori channel information available. Let I_{Ab} be the a-priori information to the left bit nodes on the edges emanating from the check nodes. Then using the Gaussian assumption, the mean of the messages from the check nodes to the bit nodes is $\mu_{Ab} = J^{-1}(I_{Ab})$. Let I_l be the channel information available to the left bit nodes with the corresponding

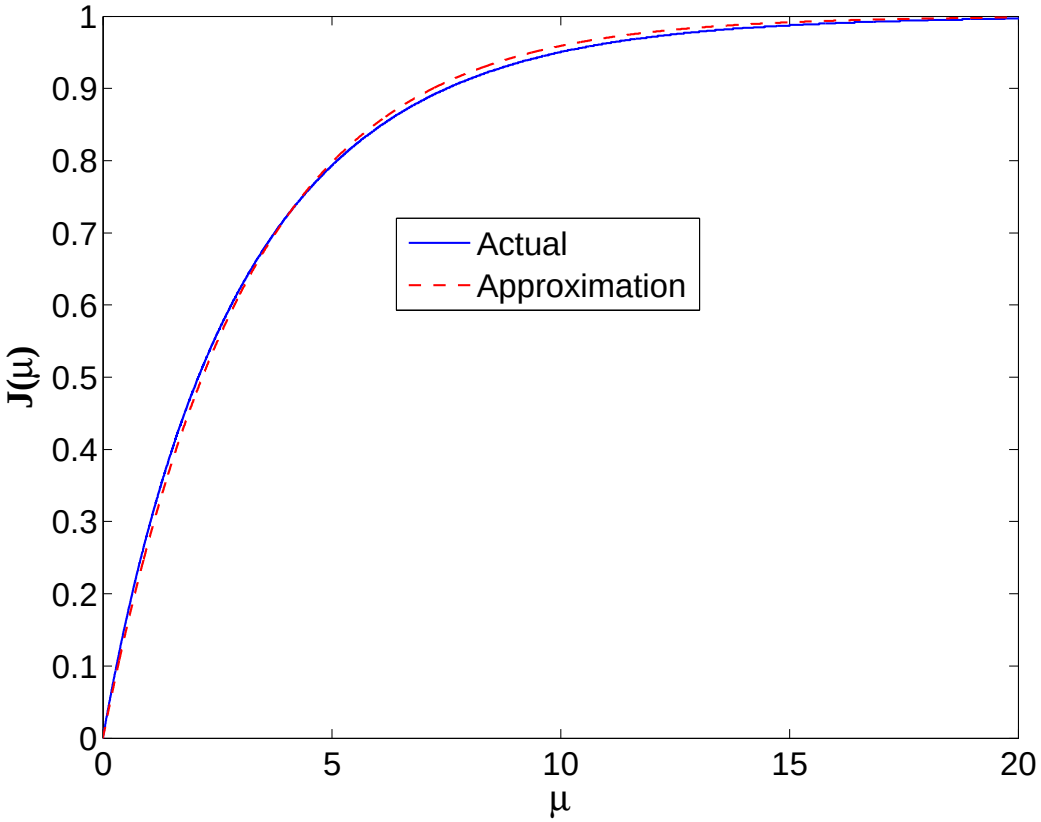


Fig. 32. Approximation of the function $J(\mu)$ by $1 - e^{-a\mu}$ with $a = 0.319$.

mean being μ_l. Then the extrinsic information from the left bit nodes is given by [65]

$$\begin{aligned} I_{Eb} &= \sum_{d=1}^{\infty} i_d J\left((d-1)\mu_{Ab} + \mu_l\right) \\ &\approx 1 - e^{-a\mu_l} \sum_{d=1}^{\infty} i_d e^{-a(d-1)\mu_{Ab}}, \end{aligned} \tag{4.17}$$

where i_d is the fraction of edges connected to degree d left bit nodes. If the average left bit node degree is γ, the degree profile $\mathcal{I}(x) = \sum_d i_d x^{d-1} \approx e^{\gamma(x-1)}$ [79, 80], with the approximation becoming exact as the length of the information sequence reaches

infinity. The extrinsic information I_{Eb} from the left bit nodes as a function of the a-priori information I_{Ab} can then be approximated as

$$\begin{aligned} I_{Eb}\left(I_{Ab}\right) &\approx 1 - e^{-a\mu_l} e^{\gamma\left(e^{-a\mu_{Ab}}-1\right)} \\ &\approx 1 - (1-I_l) e^{-\gamma I_{Ab}}, \end{aligned} \tag{4.18}$$

where we have used $e^{-a\mu_l} \approx 1 - I_l$, and $e^{-a\mu_{AB}-1} \approx -I_{Ab}$. Note that the exact value of the constant a becomes irrelevant in the approximated information transfer function.

We now take a look at the information transfer function at the check nodes. For an a-priori information I_{Ac} to the check nodes on the edges emanating from the left bit nodes, the extrinsic information at the two types of check nodes[5] is given as

$$I_{Ec}^{(i)} = \sum_{d=1}^{D} \frac{d\Omega_d^{(i)}}{\beta_i} T_d\left(I_{Ac}, I_r^{(i)}\right) \qquad i = 1, 2, \tag{4.19}$$

where β_i is the average degree of check nodes of type i, $I_r^{(i)}$ the channel information to the right bit nodes of type i, and $T_d\left(I_{Ac}, I_r\right)$ the information transfer function at a check node of degree d with I_{Ac} and I_r being the a-priori information from the left and right bit nodes, respectively. This information transfer function can be evaluated using either DDE or Monte-Carlo simulations. If α is the fraction of right bit nodes that are of type-1, and $\beta = \alpha\beta_1 + \overline{\alpha}\beta_2$ is the overall average check node degree, then the overall extrinsic information I_{Ec} from the check nodes to the left bit nodes as a function of the a-priori information I_{Ac} is given as

$$\begin{aligned} I_{Ec}\left(I_{Ac}\right) &= \frac{\alpha\beta_1}{\beta} I_{Ec}^{(1)} + \frac{\overline{\alpha}\beta_2}{\beta} I_{Ec}^{(2)} \\ &= \sum_{d=1}^{D} \frac{d}{\beta} \left[\alpha\Omega_d^{(1)} T_d\left(I_{Ac}, I_r^{(1)}\right) + \overline{\alpha}\Omega_d^{(2)} T_d\left(I_{Ac}, I_r^{(2)}\right)\right] \end{aligned} \tag{4.20}$$

[5]Each one of the two types of check nodes correspond to the two types of right bit nodes – see Fig. 30.

For successful decoding, we require that [64]

$$I_{Eb} > 1 - (1 - I_l)e^{-\gamma I_{Ec}(I_{Eb})}, \qquad \forall \quad I_{Eb} \in [0, I_{\max}],$$

where it is assumed that the precode is able to correct any decoding errors when $I_{Eb} > I_{\max}$. Expanding this convergence condition we obtain

$$\sum_{d=1}^{D} d\left[\alpha\Omega_d^{(1)} T_d\left(I, I_r^{(1)}\right) + \overline{\alpha}\Omega_d^{(2)} T_d\left(I, I_r^{(2)}\right)\right] + R_a \ln\left(\frac{1-I}{1-I_l}\right) > 0 \quad \forall \quad I \in [0, I_{\max}], \tag{4.21}$$

where $R_a = \frac{\beta}{\gamma}$ represents the overall rate of the LT code. Thus for a given α, I_l, $I_r^{(1)}$, $I_r^{(2)}$ and the node degree profiles, the maximum achievable rate R_a for the LT code is the maximum rate for which the convergence condition (4.21) is satisfied.

b. Overall code design

We now explain how the convergence condition for the general decoding setup discussed above can be utilized in designing the degree distributions for the fading relay channel. Let $\mathbf{c} = \{c_{sd}, c_{sr}, c_{rd}\}$ be the specific realizations of the three channel coefficients. Also let $\mathcal{C}_{DF}$ be the set of channel coefficients $\mathbf{c}$ for which the theoretically achievable rates of DF outperform those of CF and direct transmission. Let $\mathcal{C}_{CF}$ and $\mathcal{C}_d$ be similarly defined for the CF and direct transmission case, respectively. For some given degree distributions, let $R_{DF}(\mathbf{c})$ be the achievable LT code rates for all $\mathbf{c} \in \mathcal{C}_{DF}$ and let $R_{CF}(\mathbf{c})$ and $R_d(\mathbf{c})$ be similarly defined for the CF and the direct transmission case, respectively. One can then obtain a convergence condition for all relaying schemes using the general convergence constraint in (4.21) as follows.

Direct Transmission: For the case of direct transmission, there is only a single type of check nodes in the decoding graph, and in addition I_l is always zero. Thus, one can obtain a convergence condition for the direct transmission case by substituting

$\alpha = 1$, and $I_l = 0$ in (4.21). The check nodes follow a degree distribution $\mathbf{\Omega}^{(1)}(x)$, and the channel information to the right bit nodes is equal to the capacity on the direct link from the source to the destination, i.e., $I_r^{(1)} = C(c_{sd}^2 P)$. Thus, for all $\mathbf{c} \in \mathcal{C}_d$, successful decoding requires

$$\sum_{d=1}^{D} \Omega_d^{(1)} T_d \left(I, C(c_{sd}^2 P)\right) + R_d(\mathbf{c}) \ln(1-I) > 0 \qquad \forall \quad I \in [0, I_{\max}]. \tag{4.22}$$

Decode-Forward: Let $R_{SR}(\mathbf{c})$ be the LT code rate at which the relay is able to decode the transmission from the source. The relay decodes the source transmission over a point-to-point link, and therefore the convergence condition for decoding at the relay can be obtained on the same lines as (4.22). Thus, for all $\mathbf{c} \in \mathcal{C}_{DF}$, the rates $R_{SR}(\mathbf{c})$ should satisfy

$$\sum_{d=1}^{D} \Omega_d^{(1)} T_d \left(I, C(c_{sr}^2 P)\right) + R_{SR}(\mathbf{c}) \ln(1-I) > 0 \qquad \forall \quad I \in [0, I_{\max}]. \tag{4.23}$$

At the destination, the fraction of the relay-receive period nodes is given as $\alpha = \frac{N_1}{N} = \frac{R_{DF}(\mathbf{c})}{R_{SR}(\mathbf{c})}$. Substituting this definition of α, $I_l = 0$, $\Omega_d^{(2)} = \Omega_d^{(2)DF}$, $I_r^{(1)} = C(c_{sd}^2 P)$, and $I_r^{(2)} = C(c_{rd}^2 P)$ into (4.21) and rearranging some terms, we get the constraint on DF achievable rates as

$$\sum_{d=1}^{D} d \left[\Omega_d^{(1)} \left(\tilde{R}_{DF}(\mathbf{c}) - 1\right) T_d \left(I, C(c_{sd}^2 P)\right) + \Omega_d^{(2)DF} T_d \left(I, C(c_{rd}^2 P)\right)\right] > (1 - \tilde{R}_{DF}(\mathbf{c})) R_{SR}(\mathbf{c}) \ln(1-I) \tag{4.24}$$

$\forall\ I \in [0, I_{\max}]$, and where $\tilde{R}_{DF}(\mathbf{c}) = \frac{R_{SR}(\mathbf{c})}{R_{SR}(\mathbf{c}) - R_{DF}(\mathbf{c})}$ and $R_{SR}(\mathbf{c})$ is determined from (4.23).

Compress-Forward: Let $R_J(\mathbf{c})$ be the achievable rate for the LT code being used for DJSCC. Then for the DJSCC decoder, there is only one type of check nodes with $\Omega_d^{(1)} = \Omega_d^J$. Then for all $\mathbf{c} \in \mathcal{C}_{CF}$, the convergence condition in (4.21) for recovering

the quantized sequence $\mathbf{W}$ at the destination becomes

$$\sum_{d=1}^{D} d\Omega_d^J T_d\big(I, I_r(\mathbf{c})\big) + R^J(\mathbf{c}) \ln\left(\frac{1-I}{1-I_l(\mathbf{c})}\right) > 0 \qquad \forall \quad I \in [0, I_{\max}], \tag{4.25}$$

where $I_r(\mathbf{c}) = I\left(X_r; Y_{d2}\right)$ is the information on the relay to destination link treating the transmission from the source as interference, and $I_l(\mathbf{c}) = R_p I(W; Y_{d1})$ is the average a-priori information input to the left bit nodes. As mentioned before, the optimum choice of the half-duplexing parameter in CF is such that both $\mathbf{W}$ and $\mathbf{m}$ are decoded at the same time. This is ensured when the ratio of the length of DJSCC information sequence to that of the output codeword is $\frac{\alpha}{1-\alpha}$. In other words the optimal choice of the half-duplexing parameter is $\alpha = \frac{R^J(\mathbf{c})R_p}{1+R^J(\mathbf{c})R_p}$. Substituting this value of α and $I_l = 0$ into (4.21), we find that for all $\mathbf{c} \in \mathcal{C}_{CF}$, successful decoding of $\mathbf{m}$ for CF relaying requires

$$\begin{aligned}\sum_{d=1}^{D} d\left[R^J(\mathbf{c})R_p\Omega_d^{(1)}T_d\left(I, I_r^{(1)}(\mathbf{c})\right) + \Omega_d^{(2)CF}T_d\left(I, I_r^{(2)}(\mathbf{c})\right)\right] \\ + \left(1 + R^J(\mathbf{c})R_p\right) R_{CF}(\mathbf{c}) \ln(1-I) > 0,\end{aligned} \tag{4.26}$$

$\forall\ I \in [0, I_{\max}]$, and where $I_r^{(1)}(\mathbf{c}) = I(X_{s1}; W, Y_{d1})$, $I_r^{(2)}(\mathbf{c}) = C(c_{sd}^2 P_{s2})$ and $R^J(\mathbf{c})$ is determined from (4.25).

The average throughput over the channel ensemble is then given by

$$R_{avg} = \sum_{\mathbf{c}\in\mathcal{C}_d} p(\mathbf{c})R_d(\mathbf{c}) + \sum_{\mathbf{c}\in\mathcal{C}_{DF}} p(\mathbf{c})R_{DF}(\mathbf{c}) + \sum_{\mathbf{c}\in\mathcal{C}_{CF}} p(\mathbf{c})R_{CF}(\mathbf{c}), \tag{4.27}$$

where we have made the assumption that the fading coefficients have been discretized and $p(\mathbf{c})$ is the probability that the channel coefficients are equal to $\mathbf{c}$. Thus the optimization variables are $\{R_d(\mathbf{c})\}$, $\{R_{DF}(\mathbf{c})\}$, $\{R_{SR}(\mathbf{c})\}$, $\{R_{CF}(\mathbf{c})\}$, $\{R_J(\mathbf{c})\}$, $\left\{\Omega_d^{(1)}\right\}$, $\left\{\Omega_d^{(2)DF}\right\}$, $\left\{\Omega_d^{(2)CF}\right\}$, $\left\{\Omega_d^J\right\}$, and the objective function to maximize is R_{avg} such that the convergence conditions in (4.22), (4.23), (4.24), (4.25) and (4.26) are satisfied. In

addition, we have the trivial constraints that all degree profile coefficients should be non-negative and should sum to one. In its current form, the optimization problem is quite cumbersome, and is not convex. For our code design, we follow a simplified and possibly a sub-optimum approach, which is summarized as follows.

1. Choose the degree profile coefficients $\left\{\Omega_d^{(1)}\right\}$ to maximize $\sum_{\mathbf{c}\in\mathcal{C}} p(\mathbf{c})R_d(\mathbf{c})$ such that the convergence constraint (4.22) is satisfied for all $\mathbf{c} \in \mathcal{C}$ – the optimization can be solved using linear programming. The primary motivation for this simplification is the fact that once the coefficients $\left\{\Omega_d^{(1)}\right\}$ have been pre-designed, the CF and DF degree profiles can be designed separately.

2. **DF**: Using the pre-designed coefficients $\left\{\Omega_d^{(1)}\right\}$, find the maximum $R_{SR}(\mathbf{c})$ for all $\mathbf{c} \in \mathcal{C}_{DF}$ such that the convergence condition (4.23) is satisfied.

3. **DF**: Treating $\left\{\Omega_d^{(1)}\right\}$ and $\{R_{SR}(\mathbf{c})\}$ as constants, optimize the coefficients $\left\{\Omega_d^{(2)DF}\right\}$ to maximize $R_{avg}^{DF} = \sum_{\mathbf{c}\in\mathcal{C}_{DF}} p(\mathbf{c})R_{DF}(\mathbf{c})$ such that the convergence constraint (4.24) is satisfied. Using some algebra, the optimization can be shown to be non-linear but convex, which we solve using `CVX` [83], a tool for disciplined convex programming.

4. **CF**: Design the coefficients $\left\{\Omega_d^{J}\right\}$ to maximize the theoretical achievable rate (with practical DJSCC) $\sum_{\mathbf{c}\in\mathcal{C}_{CF}} p(\mathbf{c}) \left[\alpha^J(\mathbf{c})I_1(\mathbf{c}) + \overline{\alpha}^J(\mathbf{c})I_2(\mathbf{c})\right]$ such that the constraints in (4.25) are satisfied, and where $I_1(\mathbf{c}) = I(X_{s1}; W, Y_{d1})$, $I_2(\mathbf{c}) = C(c_{sd}^2 P_{s2})$ and $\alpha^J(\mathbf{c}) = \frac{R^J(\mathbf{c})R_p}{1+R^J(\mathbf{c})R_p}$. Once again, the optimization can be transformed to a non-linear convex problem.

5. **CF**: Treating the predesigned coefficients $\left\{\Omega_d^{(1)}\right\}$ and $\{R_J(\mathbf{c})\}$ as constants, design $\left\{\Omega_d^{(2)CF}\right\}$ to maximize the objective $\sum_{\mathbf{c}\in\mathcal{C}_{CF}} p(\mathbf{c})R_{CF}(\mathbf{c})$ such that the

convergence condition in (4.26) is satisfied. The optimization can be solved using linear programming.

E. Performance of Rateless Coded Relaying with Raptor Codes

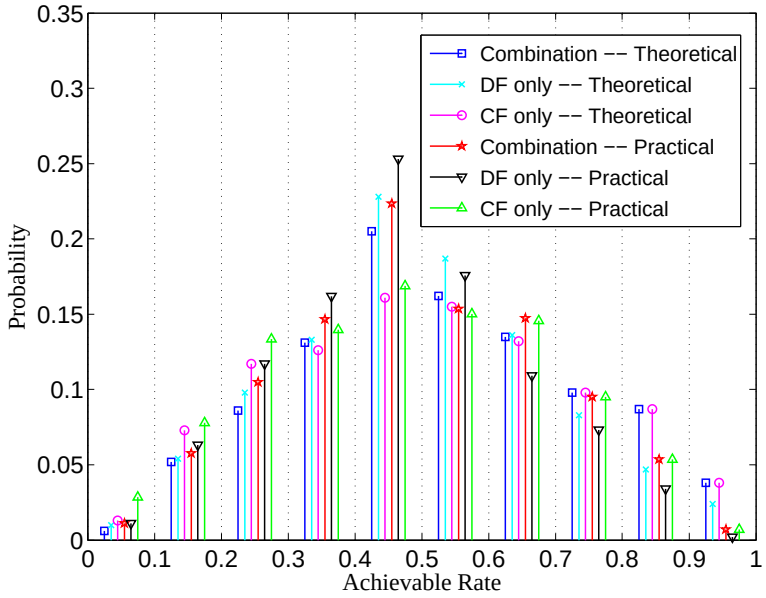
1. Practical considerations

For all Raptor codes, we use an LDPC precode of rate $R_p = 0.95$ with all bit nodes being of degree 4, and the edges to the check nodes selected uniformly (Poisson distribution). We first evaluate the performance of the designed degree distributions using DDE for asymptotically large block lengths with the number of iterations at all decoders limited to 200 – the parameters obtained from DDE are then used for finite length simulations. The performance evaluation of rateless coded DF relaying is quite straightforward. For its CF counterpart, we will try to motivate our particular choice of parameters by drawing parallels with the analysis of Section C-2. Recall that for given channel conditions, the optimum choice of α is the maximum value it can take under the condition that the quantized sequence $\mathbf{W}$ is decoded correctly at the destination. Translating this information-theoretic requirement to practice, we first find, using DDE, the maximum rate R^J for the Raptor code being used for DJSCC such that the probability of decoding error for $\mathbf{W}$ approaches zero. Using the fact that $\mathbf{W}$ and $\mathbf{m}$ should be decoded at the same time, we obtain the optimum half-duplexing parameter with DDE as $\alpha^* = \frac{R^J}{1+R^J}$. Then, following the spirit of (4.15), we find, for this optimum choice α^*, the maximum Raptor code rate corresponding to which the probability of decoding error for $\mathbf{m}$ approaches zero. For the given channel conditions, this rate is recorded as the achievable rate for rateless coded CF relaying with infinite length Raptor codes. As for the finite length Raptor codes for CF relaying, it is impossible to have some fixed N_1 which ensures, even in the

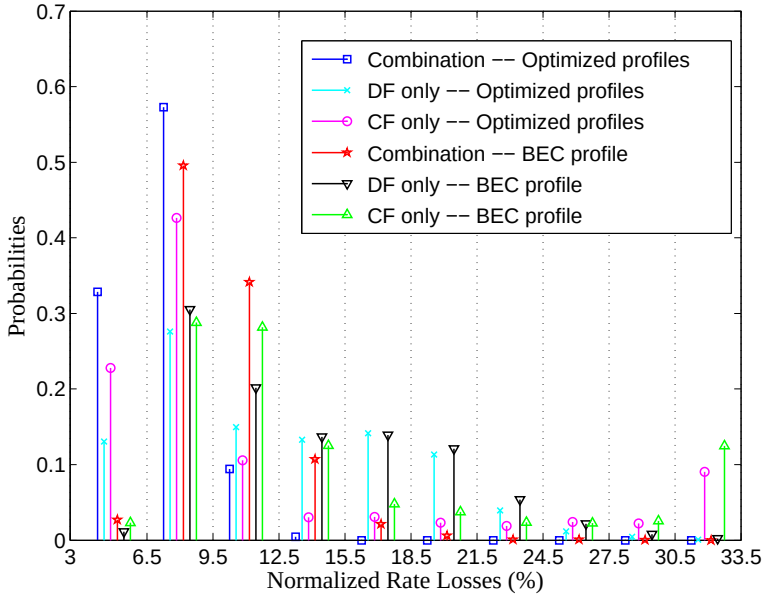
absence of fading, that the quantized sequence $\mathbf{W}$ and the source message $\mathbf{m}$ are *always* decoded at the same time. This is because the performance of a finite length Raptor code would be a function, among a host of other factors, of the exact noise vector realization − which is unknown at the nodes. Thus, as a practical solution for CF relaying, we always generate `SCH` after $N_1 = \lceil \alpha^* k/R \rceil$ symbols, where α^* is the optimum half-duplexing parameter, and R is the maximum achievable rate found with DDE. We also point out that even though the network does not have stringent delay constraints, the destination in practice cannot wait forever to decode the source information. Thus for our practical implementation, the destination stops attempting to decode after waiting for $N = 100 \times k$ symbols, and records the practical achievable rate as zero. In addition, ideally, the relay and the destination should start a new decoding attempt each time they receive a new symbol. However, that entails significant computational costs and therefore in practice we let the nodes wait for a 100 new symbols before they attempt to decode again.

2. Simulation results

We show the optimized degree distributions for $d_{sr} = 9$, $d_{rd} = 1$ and $P = 0$ dB in Table V. In Fig. 33(a), we show the probabilistic distribution of the information theoretic rates, as well as the rates obtained with the optimized Raptor code degree profiles of Table V and $k = 9500$ bits, with the maximum number of iterations at all nodes limited to 200. On the other hand, Fig. 33(b) shows the distribution of the normalized rate loss (from the information theoretic limit) of practical Raptor coded relaying with optimized degree distributions. For comparison, we also show the rate losses when using the Raptor code degree profile designed for the binary erasure channel (BEC) [79]. Some observations that can be made are:



(a)



(b)

Fig. 33. (a) Distribution of the achievable rates using finite length Raptor codes. (b) Distribution of the percent normalized rate losses using finite length Raptor codes. The distances are set at $d_{sr} = 9$, $d_{rd} = 1$, the total power $P = 0$ dB, and the information sequence length at $k = 9500$.

Table V. Designed degree distributions for $d_{sr} = 9$, and $d_{rd} = 1$ when the system power is set at $P = 0$ dB.

$\Omega^{(1)}(x)$	$0.0085x + 0.4584x^2 + 0.2511x^3 + 0.1387x^6 + 0.0329x^7 + 0.0024x^8 + 0.01x^{13} + 0.0567x^{15} + 0.0069x^{18} + 0.0338x^{46} + 0.0007x^{48}$
$\Omega^{(2)DF}(x)$	$0.5370x^2 + 0.0590x^3 + 0.1806x^4 + 0.0702x^5 + 0.0906^{11} + 0.0204x^{12} + 0.0247x^{40} + 0.0176x^{41}$
$\Omega^{(2)CF}(x)$	$0.5164x^2 + 0.1557x^3 + 0.1673x^5 + 0.0573x^7 + 0.0223x^{11} + 0.0177x^{17} + 0.0207x^{18} + 0.0426x^{29}$
$\Omega^{J}(x)$	$0.0012x + 0.3315x^2 + 0.2445x^3 + 0.2368x^5 + 0.0076x^{11} + 0.1201x^{12} + 0.0583x^{41}$

- The rate losses are (obviously) reduced when using a combination of the two schemes, instead of using the two schemes individually.

- Higher rate losses are more likely for rateless coded CF relaying as compared to DF, with the DJSCC coding losses playing a major part in the extra performance degradation. This holds true for the optimized degree profiles as well as for the BEC degree profile.

- Higher rate losses are more likely with the BEC degree profile than the optimized degree profiles, both for CF and DF relaying. However, the higher rate losses are much more pronounced for CF relaying than DF. This is because the BEC degree profiles are optimized for the case when the a-priori channel information the left bit nodes is always zero – the optimized degree profiles attempt to cater for the non-zero channel information in DJSCC decoding.

These observations can also be verified from the average throughputs shown in Table VI, which indicates that the optimized degree profiles achieve a higher average throughput than the BEC degree profile, irrespective of whether the DF or CF schemes are used individually, or in combination. In addition, CF relaying suffers a higher loss in average throughput than the DF case, with this loss being more significant for the BEC degree profile. Because of varying degree of rate losses of the two practical schemes, one would expect the probabilities of their usage (when employing a combination) to be different than that with ideal coding. As shown in Fig. 34, the probabilities of usage of the relaying schemes with Raptor codes is indeed different than those with information-theoretic codes. For example, at $d_{sr} = 9$ and $d_{rd} = 1$, the probability of usage of CF is about 3.6% less than that with ideal coding. This is explained by the fact that CF suffers higher rate losses compared to DF, as indicated in Fig. 33(b) and Table VI.

Table VI. Average throughput with finite length Raptor codes. The parameters are set at $d_{sr} = 9$, $d_{rd} = 1$, $P = 0$ dB, and $k = 9500$ bits.

	Combination	DF only	CF only
Theoretical	0.530	0.500	0.514
Optimized Degree Distributions	0.491	0.4626	0.4723
BEC Degree Distribution	0.4774	0.4583	0.4541

We also evaluate the performance of practical rateless coded relaying with Raptor codes for several other relay positions – assuming that the relay is moving along a straight line between the source and the destination. The average practical throughput along with the information theoretic bounds are shown in Fig. 35, where the practical system can be observed to lose only 0.039 (7.36%), 0.0459 (8.37%), 0.0498

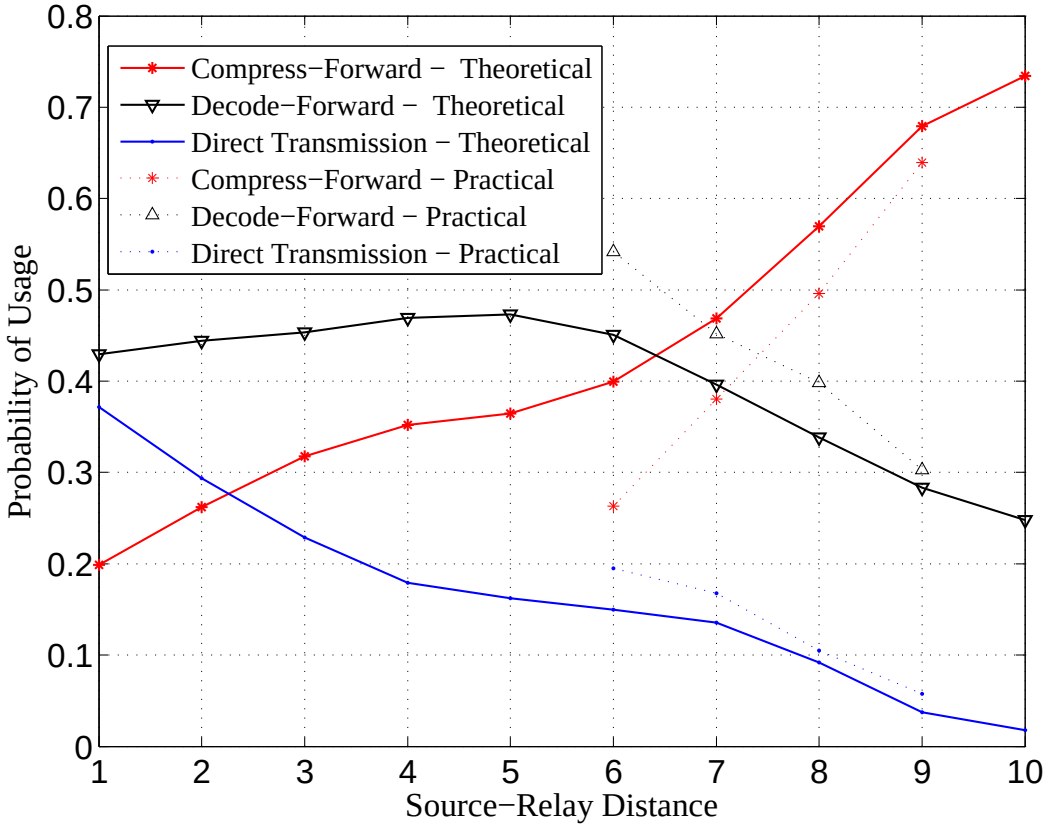


Fig. 34. Probability of usage of practical relaying schemes with optimized Raptor codes.The relay is assumed to be moving along a straight line between the source and the destination with the system power set at $P = 0$ dB.

(8.8%) and 0.0528 (9.16%) bits in average throughput from the theoretical limit when the source-to-relay distance is 9, 8, 7, and 6, respectively. On the other hand, using DDE for codes of asymptotically large block lengths, the performance loss is observed to be only 0.024 (4.52%), 0.0264 (4.8%), 0.0292 (5.17%) and 0.0299 (5.19%) bits at source-to-relay distances of 9, 8, 7, and 6, respectively. For comparison, we also plot the average throughputs obtained with the BEC profile, which is observed to be always worse than the optimized degree profiles.

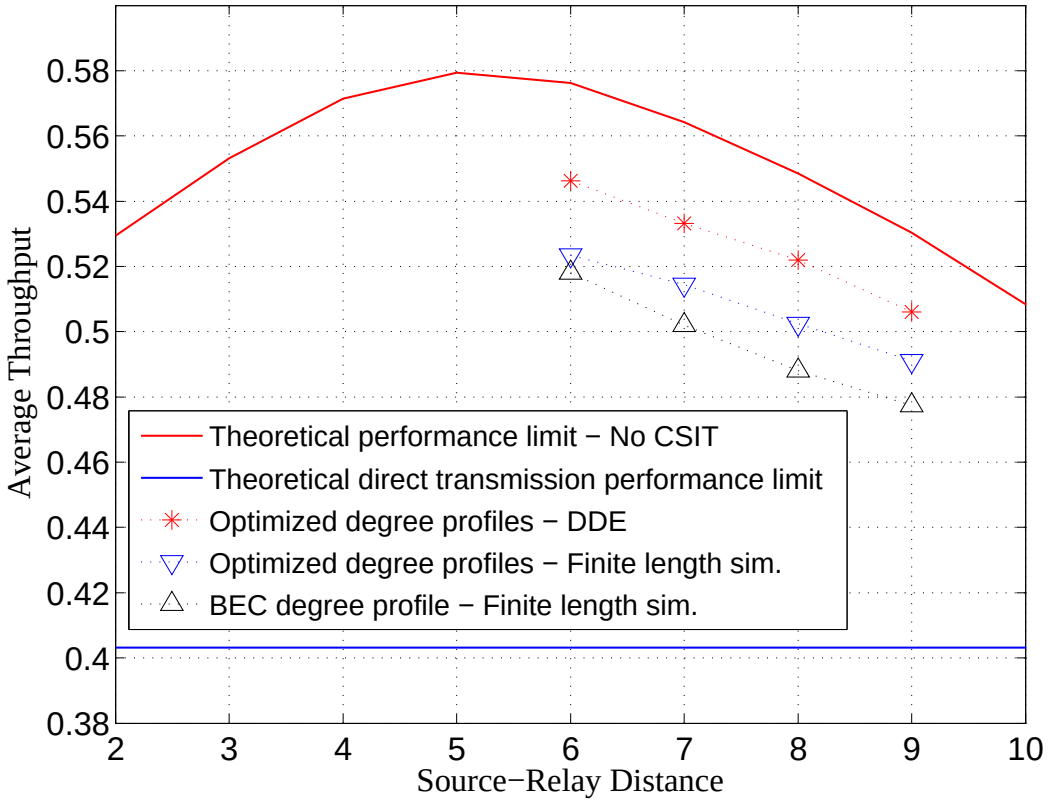


Fig. 35. Practical performance of rateless coded relaying protocol with Raptor codes. The relay is assumed to be moving along a straight line between the source and the destination with the system power set at $P = 0$ dB.

F. Summary

We have considered a half-duplex relay channel in which all links experience independent quasi-static Rayleigh fading and where the CSI is not available at the transmitters. We considered a situation where decoding delay is not a constraint and thus the source and the relay continue transmitting until successful decoding occurs at the destination. Identifying rateless coding as the natural choice, we proposed rateless coded versions of DF and CF relaying, and derived the corresponding performance limits with BPSK modulation. Since the CSI is assumed not be available at the transmitters, we proposed a protocol which allows an additional bit of feedback from the

destination, informing the source and the relay of which relaying scheme to employ, as well as indicating to the relay when to stop receiving and start transmitting. The proposed rateless coded DF and CF schemes are then put into practice using Raptor codes, which are not only used for the traditional rateless channel coding, but also for rateless DJSCC for CF relaying. The degree distributions for the Raptor codes are designed to maximize the throughput averaged over the channel ensemble, with the design formulated as a convex but non-linear optimization problem. For asymptotically large block lengths, the rateless coded relaying protocol with the optimized Raptor codes loses only $\sim 5\%$ in performance from the theoretical average throughput. With finite length simulations, the corresponding losses in average throughput are only $\sim 9\%$.

CHAPTER V

THE COOPERATIVE MULTIPLE-ACCESS CHANNEL

A. Introduction

Consider a multiple-access channel (MAC), e.g., a cellular or sensor network, where a number of nodes communicate information to a central base station or a collector node. Traditionally, nodes communicate either directly to the base station or through multi-hopping, i.e., by routing information through intermediate nodes. An alternative is to let the users cooperate when transmitting information, i.e., cooperative diversity [1, 2, 44]. Cooperation is particularly useful over multi-hopping when the channel is subject to variation due to fading or mobility, thus making the routing table quickly outdated. In this chapter, we consider the extreme of this situation where nodes have no channel knowledge. In this case, routing does not make sense, and direct transmission or cooperation are the only alternatives.

The focus of this chapter is on the energy utilized at the nodes, specifically the energy needed to transmit one bit of information. It is well known [17, 84] that for a point-to-point additive white Gaussian noise (AWGN) channel, the minimum energy per bit $\frac{E_b}{N_0}\Big|_{\min}$ approaches -1.59 dB as the bandwidth $B \to \infty$, or equivalently as the spectral efficiency (in bits/s/Hz) $R \to 0$ or the power (in Watts) $P \to 0$. This is the low power or the low SNR regime. For networks it is not known if the minimum energy per bit is approached as $R \to 0$; some results [85, 86] could indicate the contrary, i.e., the minimum energy per bit is achieved for some $R > 0$. Nevertheless, for the channel model we consider, the minimum energy per bit for the outer bounds is approached as $R \to 0$. In addition, the outer bounds and the achievable rates with the proposed cooperation methods get very close, indicating that the limit of the minimum energy

per bit as $R \to 0$ is at least close to the actual minimum energy per bit.

The motivation behind considering cooperation in the low power regime are twofold. First, in the low power regime, fading has a more significant negative impact when considering the outage capacity [87, 88], indicating the need to use cooperative diversity to overcome fading. Second, our results show that a huge energy gain is achievable through cooperation for very low spectral efficiencies. However, as soon as the spectral efficiency becomes just moderately high (say >0.5 bits/s/Hz), this gain disappears.

Nodes can operate either in full-duplex or half-duplex. In full-duplex, the nodes can receive and transmit simultaneously in the same frequency band with the transmissions possibly distinguished through different codes (code division multiple access or CDMA), whereas in half-duplex, the nodes receive and transmit in different frequency bands (frequency division multiple access or FDMA). Even though there are certain microwave techniques (e.g., circulators [89]) that make full-duplex operation possible, they currently can achieve at most 30 dB separation between transmit and receive power, which might not be enough for practical wireless networks. It is therefore appropriate to consider half-duplex operation of the nodes.It is well known that duplexing becomes irrelevant in the limit as $R \to 0$. However, as long as $R > 0$, which is always the case in practical communications, duplexing does make a difference. One way to quantify this is through the *wideband slope* [90]. We therefore use the outage capacity wideband slope for the outage capacity [91], and use this to compare CDMA and FDMA. Our results indicate that somewhat surprisingly, the difference between the two is small.

For achievable rates we consider decode-forward [92] with rateless codes [81], in conjunction with multiplexed coding. The authors in [93] considered multiplexed codes for the two-user MAC with block-Markov coding [40]. While block-Markov

codes perform slightly better than rateless codes for multiplexed coding (for superposition coding, rateless is better than block-Markov coding), they are very complicated to implement or analyze as they cause inter-block interference [94]. We therefore focus exclusively on rateless codes in this chapter. We first analyze theoretical performance, and then develop a practical coding scheme based on multiplexed Raptor codes [79] for a general N-user channel. Since the difference between CDMA and FDMA is observed to be small in the low power regime, the code design assumes that the nodes operate in half-duplex using FDMA. We simulate the practical scheme for the two- and four-user cases and find that it operates within 0.52 dB and 1.1 dB of the theoretical limit, respectively.

The rest of the chapter is organized as follows. We give our system model and define notations in Section B. Section C covers the low power regime. We derive the achievable rates for multiplexed rateless coding in Section E, followed by some numerical results in Section F. In Section G, we develop practical coding methods using Raptor codes and present simulation results. Section H provides a summary of the chapter.

B. System Model

We consider cooperative communications in a MAC with N mobile users, numbered $i = 1, \ldots, N$. The users communicate independent information to a collector (or base station) node, numbered $i = N + 1$. We denote the complex channel gain between node i and node j as c_{ij}. For notational convenience, we will often use c_i instead of $c_{i(N+1)}$ to denote the channel gain from user i to the base station. Let $X_i[n]$ and $Y_i[n]$ be the channel input and output of node i at time n, respectively. The channel can

then be modeled as

$$Y_i[n] \;=\; \sum_{j=1, j\neq i}^{N} c_{ij} X_j[n] + Z_i[n], \qquad i = 1, 2, \ldots, N+1, \tag{5.1}$$

where $Z_i[n]$ is the white Gaussian noise with power spectral density N_0. We assume that all users share the same frequency band of width B and are subject to the same power constraint, i.e., $\mathbb{E}[X_i^2] \leq P$. In addition, for notational simplicity we assume without loss of generality that $N_0 = 1$. Consequently we will denote the ratio $\frac{E_b}{N_0}$ as E_b in the sequel and refer to it merely as the energy required to transmit one bit of information. All channel gains c_{ij} are assumed to experience i.i.d. block *flat* fading. Furthermore, the nodes are assumed to have no channel state information, except as required for decoding. What a node needs to know is if it has decoded a packet correctly, so that it can forward it, which could be ensured for example by error-detection coding. A reasonable performance measure is therefore the *outage capacity*. The above model is relevant in a number of real-world scenarios.

- **Rayleigh fading**: Consider a set of nodes scattered in a small area. In that case, path loss is a minor factor, while fading is the dominant factor.

- **Mobile nodes**: Consider a set of nodes that move around. Their path loss coefficients, shadowing, and fading coefficients will be varying. If they move around rapidly or transmit only in short bursts, it does not make sense to build a routing tree structure. In this case, the outage probability is determined by the spatial distribution of the nodes.

- **A combination of the above two**.

It should be mentioned that our theoretical analysis in Sections C and D is not limited to this model. However, the achievable rates presented in Section E and the code

designs in Section G are aimed at the above scenario, in which the nodes transmit their information without using any knowledge about the network and forward whatever other packets they can decode. Our code designs will also work if nodes are in fixed positions and they experience mainly path loss, but in that case it is probably more efficient to estimate the channel, build a routing table, and use multi-hop communications.

As mentioned before, we consider two modes of node cooperation. In the first mode, the nodes are assumed to operate in full-duplex, i.e., they can transmit and receive at the same time on the same frequency band. In that case all nodes transmit simultaneously, using different codes, and decoding is done jointly, i.e., with multiuser detection. We refer to this mode as CDMA. It should be pointed out here that we use the term CDMA in the context it was used in [17], on pp. 547 – it should not be confused with DS-CDMA. In the second mode, the nodes can operate only in half-duplex. Half-duplexing is achieved by dividing the total frequency band into N subbands of equal bandwidth with node i transmitting in subband i and listening on the remaining $N-1$ subbands. A relevant model in this case could be to let subbands fade independently. Both our theoretical analysis and code designs work just as well for this scenario with independent fading. However, to compare with CDMA, we restrict our analysis to the flat fading case. In addition, when generating numerical and simulation results, we assume channel reciprocity for both CDMA and FDMA, i.e. $c_{ij} = c_{ji}$. Whereas this assumption makes sense for CDMA, it might not be valid for FDMA. However, we point out that all achievable rates we derive are independent of this assumption. We make this assumption only to provide a fair comparison of FDMA with CDMA when we generate the numerical results.

In order to clarify concepts related to the MAC, consider first a point-to-point block fading AWGN channel. Let B be the bandwidth of the channel. If the fading

coefficient c is fixed and known then

- The spectral efficiency[1] R in *bits per complex sample* is said to be achievable if there exists a sequence of $(2^{n(R-\epsilon)}, n)$ codes with asymptotically zero error probability for any $\epsilon > 0$. For the Gaussian channel, the spectral efficiency $R = \log_2\left(1 + |c|^2 \frac{P}{B}\right)$ is achievable and is of course the capacity.

- The rate of the transmission is BR in *bits/second*.

- The transmitted energy (in dB) per information bit as a function of the channel coefficient c and the spectral efficiency R is given as

$$E_b(c, R) = 10 \log_{10}\left(\frac{P}{BR}\right). \tag{5.2}$$

Since nodes have no channel state information at the time of transmission, it is reasonable to consider the outage rate/capacity as a performance measure. The spectral efficiency R (for a given power P) is said to be achievable with outage probability p if for all $\delta, \epsilon > 0$ there exists a $(2^{n(R-\epsilon)}, n)$ code with asymptotically zero error probability except on a set of channel states with probability $p + \delta$. For the point-to-point AWGN channel, let

$$R(p) \;=\; \max\left\{ \mathrm{R} \,\middle|\, \Pr\left(\mathrm{R} > \log_2\left(1 + |c|^2 \frac{P}{B}\right)\right) \leq p \right\}. \tag{5.3}$$

Then is is clear that $R(p)$ is the maximum achievable spectral efficiency with outage probability p. Thus, the outage spectral efficiency can be found by calculating the distribution of $\log_2\left(1 + |c|^2 \frac{P}{B}\right)$(as a function of c), and choosing $R(p)$ as the p percentile. On the other hand, for a given transmission spectral efficiency R and an

[1]This is often called the rate, but in accordance with [90] we will call it the spectral efficiency.

outage probability p, the *outage* energy per bit $E_b(p,R)$[2] can be found as

$$E_b(p,R) = \min\left\{ e \;\middle|\; \Pr\left(e < E_b(c,R)\right) \le p \right\}. \tag{5.4}$$

Consider now the N-user MAC. We assume that each user $i, 1 \le i \le N$, has the same power constraint P and needs to send independent information with the same spectral efficiency $R_i = R$ to the collector with a delay requirement that is short compared to the coherence time of the channel (the results can be easily generalized to nodes with different rate and power requirements). In other words, the channel coefficients remain constant during a block of transmission. An outage event is declared if *at least one* of the users cannot communicate at the target spectral efficiency R (similar to the *individual* outage capacity in [95]). Let $\mathbf{R}(\mathbf{c})$ be an achievable spectral efficiency region for a specific set of channel coefficients $\mathbf{c}$. Define

$$R(p) \;=\; \max\left\{ \mathrm{R} \;\middle|\; \Pr\left((\mathrm{R},\mathrm{R},\ldots,\mathrm{R}) \notin \mathbf{R}(\mathbf{c})\right) \le p \right\}. \tag{5.5}$$

Then it seems reasonable that $R(p)$ should be achievable with outage probability p. Indeed, while this is not obvious (cf. [81]), it turns out to be true for the coding schemes we consider in this chapter. If $\mathbf{R}(\mathbf{c})$ is an outer bound on the achievable spectral efficiency, it is clear, on the other hand, that (5.5) gives an outer bound on the achievable outage spectral efficiency.

C. The Low Power Regime

In order to extend node and network life, a critical consideration in the design of the communication systems is power efficiency. Specifically, we want to minimize the

[2]With a slight abuse of notation, we use $E_b(c,R)$ to denote the energy per bit for a fixed channel, while $E_b(p,R)$ is the *outage* energy per bit.

energy required for transmitting each bit of information. In a point-to-point AWGN channel, it is known [17, 90] that the rate $B\log_2\left(1+|c|^2\frac{P}{B}\right)$ is monotonically increasing with B, and the minimum energy required to transmit one bit of information is therefore reached as $B\to\infty$, which leads to the spectral efficiency $R\to 0$. For networks in general, e.g., the relay network [85], it is not known if the rate is monotonic in B or if the minimum energy per bit is reached for $B\to\infty$. However, all achievable rates, as well as outer bounds considered in this chapter are monotonic in B, hence the minimum energy per bit is indeed reached as $B\to\infty$. We denote this limit of E_b when $B\to\infty$ as $E_{b,min}$.

The investigation of $E_{b,\min}$ has been concentrated on the ergodic capacity [90, 96, 97] for point-to-point channels and later extended to relay channels [85, 86, 88, 98] for ergodic or outage capacity. In [86], the authors derived bounds on $E_{b,\min}$ in AWGN relay channels. In [85], the authors considered the achievable $E_{b,\min}$ for ergodic capacity in fading relay channels. It was later shown in [88, 98] that bursty amplify-forward achieves the ϵ-outage capacity and consequently the ϵ-outage capacity per unit cost when the outage probability ϵ is arbitrarily small with or without channel state information at the receivers. In this chapter we focus on a finite, fixed outage probability.

As mentioned earlier, we assume that all users transmit at the same spectral efficiency R. The energy per bit $E_b(\mathbf{c},R)$ required to achieve a spectral efficiency R for a given channel state $\mathbf{c}$ is a random variable, which we will assume to have a continuous distribution $f_E(x;R)$. We define the minimum energy per bit as a function of the channel coefficients as $E_{b,\min}(\mathbf{c})=E_b(\mathbf{c},0)$. The outage energy $E_b(p,R)$ can then be written as

$$E_b(p,R) \;=\; \min\left\{e\,\middle|\int_{x>e} f_E(x;R)dx\le p\right\}. \tag{5.6}$$

We now have

$$E_{b,\min}(p) = \lim_{R\to 0} \min\left\{ e \left| \int_{x>e} f_E(x;R)dx \le p \right.\right\}$$

$$= \min\left\{ e \left| \int_{x>e} \lim_{R\to 0} f_E(x;R)dx \le p \right.\right\}, \quad (5.7)$$

assuming the distribution $f_E(x;R)$ is sufficiently nice. Therefore we can also write

$$E_{b,\min}(p) = \min\left\{ e \left| \int_{E_{b,\min}(\mathbf{c})>e} f(\mathbf{c})d\mathbf{c} \le p \right.\right\}. \quad (5.8)$$

The expression for $E_{b,\min}(p)$ in (5.8) simplifies performance analysis, as we just have to find the function $E_b(\mathbf{c},0) = E_{b,\min}(\mathbf{c})$. For simple networks, we can further solve the inequality $E_{b,min}(\mathbf{c}) > e$ analytically. In other cases we have to calculate $\int_{E_{b,\min}(\mathbf{c})>e} f(\mathbf{c})d\mathbf{c}$ using Monte Carlo simulations instead.

Verdu pointed out in [90] that the limit of E_b as $B \to \infty$ is not necessarily a good indicator of performance for large, but finite, B. For example, in the context of this chapter, FDMA and CDMA have the same $E_{b,\min}$, but their spectral usage is very different. Verdu therefore introduced the *wideband slope* defined as

$$\mathcal{S}_0 = \lim_{E_b\to E_{b,\min}} \frac{R(E_b)}{E_b - E_{b,min}} 10\log_{10} 2 \quad (5.9)$$

$$= \lim_{B\to\infty} \frac{\partial R(E_b)}{\partial E_b} 10\log_{10} 2, \quad (5.10)$$

where E_b is given in dB. The wideband slope for outage rate is not quite as straightforward to find as $E_{b,\min}(p)$, but it can be obtained by using Theorem 1 in [91]. The theorem has been reproduced below for the reader's convenience.

Theorem 1 *Assume that the fading distribution is continuous and that its pdf is*

continuously differentiable. Let

$$\nabla E_{b,\min}(\mathbf{c}) = \left(\frac{\partial E_{b,\min}(\mathbf{c})}{\partial c_1}, \ldots, \frac{\partial E_{b,\min}(\mathbf{c})}{\partial c_M} \right) \tag{5.11}$$

and $\mathcal{C}(p) = \{\mathbf{c} | E_{b,\min}(\mathbf{c}) = E_{b,\min}(p)\}$. *Assume that* $\mathcal{C}(p)$ *is a compact, differentiable manifold. Then*

$$\begin{aligned} \mathcal{S}_0^{-1}(p) &= \left(\int_{\mathcal{C}(p)} \frac{1}{\|\nabla E_{b,\min}(\mathbf{c})\| \mathcal{S}_0(\mathbf{c})} f(\mathbf{c}) d\mathbf{c} \right) \times \\ &\quad \left(\int_{\mathcal{C}(p)} \frac{1}{\|\nabla E_{b,\min}(\mathbf{c})\|} f(\mathbf{c}) d\mathbf{c} \right)^{-1}, \end{aligned} \tag{5.12}$$

where the integrals denote integration over the $N-1$ *dimensional manifold* $\mathcal{C}(p)$ *with induced measure.*

Alternatively,

$$\begin{aligned} \mathcal{S}_0^{-1}(p) &= \lim_{\delta \to 0} \left(\int_{|E_{b,\min}(\mathbf{c}) - E_{b,\min}(p)| < \delta} \frac{1}{\|\nabla E_{b,\min}(\mathbf{c})\| \mathcal{S}_0(\mathbf{c})} f(\mathbf{c}) d\mathbf{c} \right) \times \\ &\quad \left(\int_{|E_{b,\min}(\mathbf{c}) - E_{b,\min}(p)| < \delta} \frac{1}{\|\nabla E_{b,\min}(\mathbf{c})\|} f(\mathbf{c}) d\mathbf{c} \right)^{-1}. \end{aligned} \tag{5.13}$$

The reader is referred to [91] for a detailed proof. Notice that if $E_{b,\min}$ only depends on $|c_n|^2$, the derivative in (5.11) can be computed with respect to $|c_n|^2$. Also note that if the wideband slope $\mathcal{S}_0$ is independent of the channel coefficients, the outage wideband slope is given as $\mathcal{S}_0(p) = \mathcal{S}_0$, and is therefore always independent of the outage probability.

D. Bounds for Cooperation

In this section, we will compare FDMA and CDMA by considering some simple bounds for cooperation and evaluating the corresponding $E_{b,\min}(p)$ and the outage wideband slopes. In the following, we consider the lower and upper bounds for the

two schemes separately.

1. Lower (No Cooperation) bounds

A simple lower bound on cooperation is of course provided by the no cooperation case[3]. For a fixed set of channel coefficients $\mathbf{c}$, the capacity region for the CDMA case is well known [17]. The minimum energy per bit is obtained as

$$\begin{aligned} E_{b,\min}(\mathbf{c}) &= 10\log_{10}\left(\frac{\ln 2}{\min_{i=1\ldots N}\left\{|c_i|^2\right\}}\right) \\ &= -1.59 - 10\log_{10}\left(\min_{i=1\ldots N}\left\{|c_i|^2\right\}\right). \end{aligned} \tag{5.14}$$

We can now use (5.8) to find the outage minimum energy (in dB) as

$$E_{b,\min}(p) = -1.59 - 10\log_{10}\left(\max\left\{c\left|\Pr\left(\min_{i=1\ldots N}\left\{|c_i|^2\right\} < c\right) \le p\right.\right\}\right). \tag{5.15}$$

This expression can be evaluated analytically, since it depends on the distribution of the minimum of N χ-squared random variables. Using Theorem 9 in [90], we can find the wideband slope as $\mathcal{S}_0 = 2$. Since this is independent of the channel, it is also the outage wideband slope by Theorem 1.

With FDMA, the achievable spectral efficiency for the no cooperation case is given by

$$R = \frac{1}{N}\min_{i=1\ldots N}\log\left(1+|c_i|^2 N\mathrm{SNR}\right). \tag{5.16}$$

This gives the same $E_{b,min}$ as that in (5.15) for the CDMA case. However, this does not mean that FDMA and CDMA are equivalent in the low power regime since their

[3]We will consider tighter lower bounds in the form of achievable rates with cooperation in the next section

wideband slopes could be very different. Indeed, using Theorem 9 in [90], the wideband slope for FDMA can be evaluated as $\mathcal{S}_0 = \frac{2}{N}$, which is also the outage wideband slope. Thus, while FDMA with no cooperation achieves the same minimum energy as CDMA, FDMA operation needs approximately N times as much bandwidth, and is therefore very inefficient.

2. Upper bounds

An outer bound for cooperative communications can be obtained by assuming that all nodes know all messages. If the nodes use CDMA, this outer bound is given by

$$R = \frac{1}{N}\log\left(1+\sum_{i=1}^{N}|c_i|^2\mathrm{SNR}\right). \tag{5.17}$$

Similarly for FDMA, this outer bound can be evaluated as

$$R = \frac{1}{N^2}\sum_{i=1}^{N}\log\left(1+|c_i|^2N\mathrm{SNR}\right). \tag{5.18}$$

For both schemes we get

$$E_{b,\min}(\mathbf{c}) = -1.59 - 10\log_{10}\left(\frac{1}{N}\sum_{i=1}^{N}|c_i|^2\right) \tag{5.19}$$

and the corresponding outage energy per bit is

$$E_{b,\min}(p) = -1.59 - 10\log_{10}\left(\max\left\{c\left|\Pr\left(\frac{1}{N}\sum_{i=1}^{N}|c_i|^2<c\right)\le p\right.\right\}\right), \tag{5.20}$$

which can be evaluated analytically. For CDMA, the wideband slope associated with the outer bound of (5.18) can be evaluated as $\mathcal{S}_0(\mathbf{c}) = \frac{2}{N}$ using Theorem 9 in [90]. Using the same Theorem for the FDMA bound in (5.18) gives

$$\mathcal{S}_0(\mathbf{c}) = \frac{2\left(\sum_{i=1}^{N}|c_i|^2\right)^2}{N^2\sum_{i=1}^{N}|c_i|^4}. \tag{5.21}$$

Since this depends on $\mathbf{c}$, we cannot directly obtain the wideband slope for outage capacity. We can instead use Theorem 1. Notice that the rate only depends on $|c_i|^2$, so we can do all calculations with respect to these coefficients, leading to

$$\|\nabla E_{b,min}(\mathbf{c})\| = \frac{10}{\ln 10}\left(\sum_{i=1}^{N}|c_i|^2\right)^{-1}. \quad (5.22)$$

Since the integration in (5.12) is over the region specified by $E_{b,\min}(\mathbf{c}) = E_{b,\min}(p)$,

$$\sum_{i=1}^{N}|c_i|^2 = \frac{N\ln 2}{10^{E_b(p)/10}} \triangleq \kappa \quad (5.23)$$

is fixed. This implies that $\|\nabla E_{b,min}(\mathbf{c})\|$ is also a constant. We can therefore write the wideband slope as

$$\mathcal{S}_0^{-1}(p) = \int_{\sum_{i=1}^{N}|c_i|^2=\kappa} \frac{N^2\sum_{i=1}^{N}|c_i|^4}{2\left(\sum_{i=1}^{N}|c_i|^2\right)^2} f(\mathbf{c})d\mathbf{c}\left(\int_{\sum_{i=1}^{N}|c_i|^2=\kappa} f(\mathbf{c})d\mathbf{c}\right)^{-1}. \quad (5.24)$$

If the distribution $f(\mathbf{c})$ is circular symmetric, this value is independent of κ. This means that the wideband slope is independent of the outage probability p, and can be re-written as

$$\mathcal{S}_0^{-1}(p) = \int_{\mathbb{R}^N} \frac{N^2\sum_{i=1}^{N}|c_i|^4}{2\left(\sum_{i=1}^{N}|c_i|^2\right)^2} f(\mathbf{c})d\mathbf{c}. \quad (5.25)$$

While it is difficult to evaluate this integral analytically, it can easily be evaluated by Monte-Carlo integration. Fig. 36 shows the wideband slope for Rayleigh fading. While the wideband slopes for CDMA and FDMA considered above are associated with the upper bounds, we will later see that we can achieve rates very close to the upper bounds, and we will therefore briefly discuss these results here.

As indicated in Section F, FDMA and CDMA give the same big gain in minimum E_b. The wideband slopes are not very different either, indicating that FDMA has little loss compared to CDMA. This is in stark contrast to the no cooperation case,

where FDMA is significantly inferior to CDMA. On the other hand, both cooperation methods have much smaller slope than non-cooperative CDMA. This indicates that the gain from cooperation quickly evaporates when leaving the very low power regime. These observations, verified by numerical results, will be discussed in more detail in Section F.

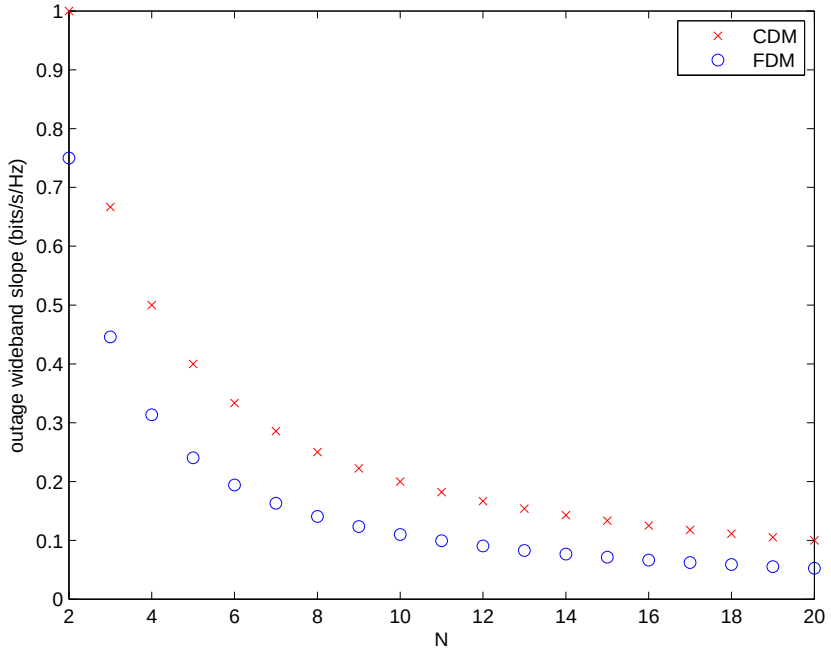


Fig. 36. The wideband slope (for different number of users) based on the rate upper bound for cooperation under Rayleigh fading.

E. Achievable Rates using Multiplexed Rateless Codes

In this section, we will derive achievable rates for multiplexed rateless coding. We explain the main ideas for the two-user MAC before generalizing to the N-user case.

1. Cooperation methods

There are three basic cooperation strategies: decode-forward, amplify-forward, and compress-forward [40, 43, 61]. The latter two do not make much sense in the low SNR regime since at low power almost all of the received signals are pure noise, and forwarding this is not sensible. In fact for amplify-forward, $E_{b,min}$ is reached at finite SNR [85]. We will therefore only consider decode-forward. It is shown in [98, 88] that bursty amplify-forward scheme achieves the outage capacity in the low power regime when the outage probability $p \to 0$. However, for a finite outage probability, bursty amplify-forward is not advantageous over decode-forward in the low power regime [94]. For example, for a simple three-node relay channel with i.i.d. Rayleigh fading on all links, the E_b performance for outage probability 10^{-2} is shown in Fig. 37. It is seen that in the low power regime (when data rate is very low), decode-forward performs better than bursty amplify-forward.

In decode-forward, a node needs to re-encode its decoded messages from the partner together with its own message. A commonly used coding scheme is *superposition coding*. Suppose that the transmitter wants to transmit two messages $w_1 \in \{1, \dots 2^{nR_1}\}$ and $w_2 \in \{1, \dots 2^{nR_2}\}$. The transmitter then splits its transmission power P between the messages to be sent and each of the messages is encoded by an independent codebook. The resulting signal is the superposition of both corresponding codewords. Thus the transmitter has one codebook $\mathbb{C}_1$ for w_1 and encodes w_1 to $\mathbf{X}_1(w_1)$ and another independent codebook $\mathbb{C}_2$ for w_2 and encodes w_2 to $\mathbf{X}_2(w_2)$. Both $\mathbf{X}_1$ and $\mathbf{X}_2$ have average power P. The resulting signal is then $\mathbf{X}(w_1, w_2) = \sqrt{\alpha}\mathbf{X}_1(w_1) + \sqrt{1-\alpha}\mathbf{X}_2(w_2)$, where $\alpha \in [0, 1]$. It is seen that if the receiver has the knowledge of w_1 (or w_2), part of the transmission power is wasted. In a two-user MAC, if w_1 is a user's own message and w_2 is the message it decoded from

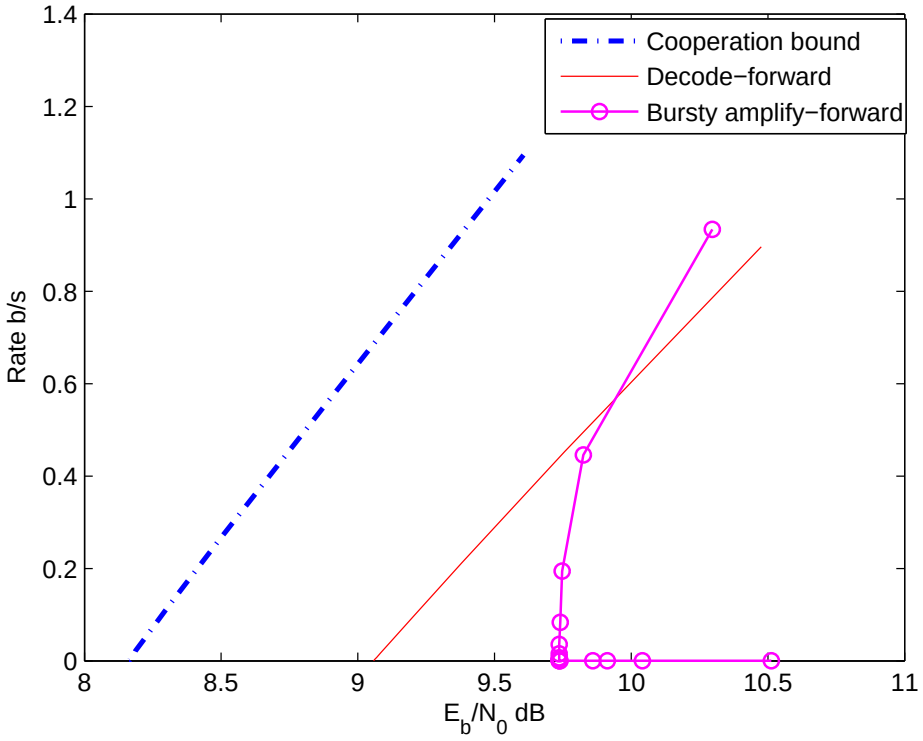


Fig. 37. $\frac{E_b}{N_0}$ versus rate performance when the outage probability is 10^{-2} for a three--node relay channel with i.i.d. Rayleigh fading.

the transmission of its partner, from the perspective of the partner, part of the transmission power $(1-\alpha)P$ is wasted and thus the likelihood of cooperation is reduced. Increasing α will increase the likelihood of cooperation but decrease the cooperation efficiency, which depends on how much power is assigned for the transmission of the partner's messages by a user.

A coding scheme which circumvents the above problem is called *multiplexed coding* [93]. The idea is as follows: A codeword can convey different kinds of information depending on the side information the receiver has. The transmitter makes a table with 2^{nR_1} rows and 2^{nR_2} columns, and assigns a random (Gaussian) code to each entry in the table, denoted as $\mathbf{X}(w_1, w_2)$. A receiver can decode both w_1 and w_2 if the

channel capacity $C > R_1 + R_2$. If it knows w_1, however, it can decode w_2 if $C > R_2$ simply by searching only the row corresponding to w_1, and similarly if it knows w_2 it can decode w_1 if $C > R_1$. Multiplexed coding applies to both binary and non-binary codes and can be extended to more messages. A related scheme is in [99], where the authors XOR binary codes. However, this cannot be used in the kind of networks we consider, as a receiver has to be able to decode both messages if neither is known.

Multiplexed coding and superposition coding can be implemented using block Markov [100] or rateless codes [101]. In [93], block Markov coding for the two-user case was explored. However, block Markov coding is not easy to scale to more than two users due to inter-block interference [94], and therefore we only consider rateless codes.

2. The two-user case

Consider first a point-to-point link, and suppose that the channel gain c is unknown. A $(n, 2^{nR})$ code can be constructed so that the destination can decode the message with low probability of error after n_1 symbols if

$$n_1 \log(1 + |c|^2 \mathrm{SNR}) \geq nR, \tag{5.26}$$

where $\mathrm{SNR} = \frac{P}{B}$. Define $t = \log(1 + |c|^2 \mathrm{SNR})^{-1}$. Then if $t^{-1} \geq R$, the destination can decode the message without outage; if $t^{-1} < R$, there is an outage. Furthermore, if n tends towards infinity, then the destination can decode after a fraction tR of the whole block. We can therefore view t as the time it takes for the destination to decode the message. This is the principle of rateless coding, and in the following we will apply this to the two-user MAC. Since this is a compound channel, going from the rate for a specific channel to outage rate requires a formal proof as in [81]. However, we will keep the discussion in this chapter more informal – we refer the

reader to [11] for a formal proof. In addition, for better explaining the basic concepts behind the calculation of achievable rates for the two user case, we will assume that $c_{12} = c_{21}$. The general N-user case without channel reciprocity is discussed in the next subsection.

Consider first the CDMA case. If $|c_{12}| \leq \min\{|c_1|, |c_2|\}$, the two users do not cooperate. From the achievable rate region of a non-cooperative MAC, we get

$$\begin{aligned} R &\leq \log\left(1 + |c_1|^2 \mathrm{SNR}\right) \\ R &\leq \log\left(1 + |c_2|^2 \mathrm{SNR}\right) \\ R &\leq \frac{1}{2}\log\left(1 + (|c_1|^2 + |c_2|^2)\mathrm{SNR}\right) \end{aligned} \tag{5.27}$$

On the other hand, if $|c_{12}| > \min\{|c_1|, |c_2|\}$, the users cooperate. Because of the symmetry of the channel, we can assume without loss of generality that $|c_1| < |c_2|$. The two users can decode each other after a duration of

$$t_c = \frac{1}{\log(1 + |c_{12}|^2 \mathrm{SNR})}. \tag{5.28}$$

After decoding, user 2 transmits messages 1 and 2 using a multiplexed codebook (as does user 1). If the transmission continues for a total time t, the collector can decode message 1 if

$$R \leq (t - t_c)\log(1 + (|c_1|^2 + |c_2|^2)\mathrm{SNR}) + t_c \log(1 + |c_1|^2 \mathrm{SNR}). \tag{5.29}$$

There is a similar condition for decoding message 2 at the collector, but this bound is larger. Finally the condition for joint decoding is given as

$$R \leq \frac{1}{2}\log\left(1 + (|c_1|^2 + |c_2|^2)\mathrm{SNR}\right). \tag{5.30}$$

If we define

$$\begin{aligned}
t_c &= \frac{1}{\log(1+|c_{12}|^2\mathrm{SNR})}, \\
t_1 &= \frac{1}{\log(1+|c_1|^2\mathrm{SNR})}, \\
t_2 &= \frac{1}{\log(1+|c_2|^2\mathrm{SNR})}, \\
t_{12} &= \frac{2}{\log(1+(|c_1|^2+|c_2|^2)\mathrm{SNR})}, \\
\tilde{t}_1 &= \frac{1-t_c\log(1+|c_1|^2\mathrm{SNR})}{\log(1+(|c_1|^2+|c_2|^2)\mathrm{SNR})}+t_c, \\
\tilde{t}_2 &= \frac{1-t_c\log(1+|c_2|^2\mathrm{SNR})}{\log(1+(|c_1|^2+|c_2|^2)\mathrm{SNR})}+t_c,
\end{aligned} \tag{5.31}$$

then we can rewrite the achievable spectral efficiency as

$$R = \frac{1}{t}, \tag{5.32}$$

$$t = \max\{\min\{t_1,\tilde{t}_1\},\min\{t_2,\tilde{t}_2\},t_{12}\}. \tag{5.33}$$

Similarly, for FDMA, the spectral efficiency can also be evaluated using (5.32) and (5.33) but with the following definitions

$$\begin{aligned}
t_c &= \frac{2}{\log(1+|c_{12}|^2 2\mathrm{SNR})}, \\
t_1 &= \frac{2}{\log(1+|c_1|^2 2\mathrm{SNR})}, \\
t_2 &= \frac{2}{\log(1+|c_2|^2 2\mathrm{SNR})}, \\
t_{12} &= \frac{4}{\log(1+|c_1|^2 2\mathrm{SNR})+\log(1+|c_2|^2 2\mathrm{SNR})}, \\
\tilde{t}_1 &= \frac{1-\frac{t_c}{2}\log(1+|c_1|^2 2\mathrm{SNR})}{\frac{1}{2}\log(1+|c_1|^2 2\mathrm{SNR})+\frac{1}{2}\log(1+|c_2|^2 2\mathrm{SNR})}+t_c, \\
\tilde{t}_2 &= \frac{1-t_c\log(1+|c_2|^2\mathrm{SNR})}{\frac{1}{2}\log(1+|c_1|^2 2\mathrm{SNR})+\frac{1}{2}\log(1+|c_2|^2 2\mathrm{SNR})}+t_c.
\end{aligned} \tag{5.34}$$

An intuitive interpretation of (5.31) is shown in Fig. 38. Each one of the three

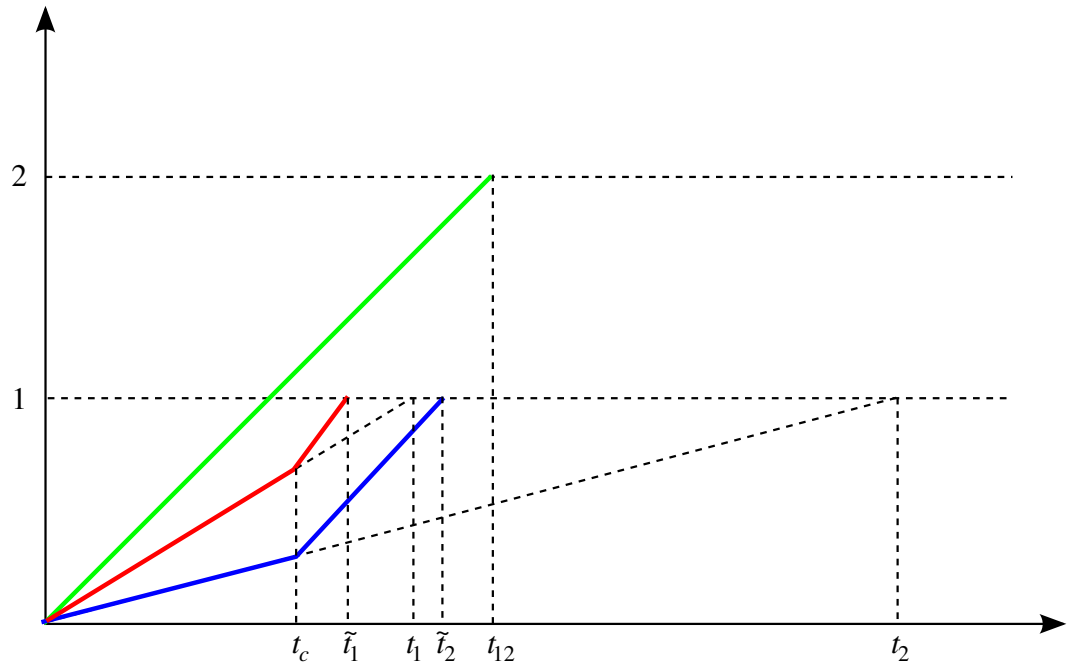


Fig. 38. Bucket filling interpretation of rateless multiplexed coding.

rate constraints of a MAC can be thought of as an empty bucket of unit volume. Two of these buckets $B^1$and B^2 correspond to the individual rate constraints (5.29) of each user and are to be filled up by the individual information of user 1 and 2, respectively. The third bucket B^{12} corresponds to the sum-rate constraint (5.30) and is to be filled up by the joint information of both users. A user's information is decodable at the collector if the two buckets associated with this user are both full. Each user also has a unit volume bucket to be filled up by the information of the other user. The rate to fill up the users' bucket is $\log(1 + |c_{12}|^2\text{SNR})$. For the collector's buckets, initially, the rate of filling up the bucket B^i, $i = 1, 2$, is $\log(1 + |c_i|^2\text{SNR})$ and for bucket B^{12}, $1/2 \log(1 + (|c_1|^2 + |c_2|^2)\text{SNR})$. If after $n_1 < n$ symbols (after time t_c), the two users' buckets are full, the two users are able to decode each other and can

therefore start transmitting cooperatively. At this point, the fill rate of $B^i, i = 1, 2$, should be updated to $\log(1+(|c_1|^2+|c_2|^2)\text{SNR})$ but the rate for B^{12} remains the same (cooperation does not give any gain for sum-rate). A rate of $R = 1/t$ is achievable if after a time t all three buckets of the collector are full. A similar strategy with different fill rates can be used to calculate the spectral efficiency for the FDMA case.

To get the achievable $E_{b,\min}$, the easiest way is to redefine quantities in terms of rate instead of spectral efficiency. So, in (5.31) (equivalently (5.34)) we replace SNR with $\frac{P}{B}$ and $\log(\cdot)$ with $B\log(\cdot)$. In the limit as $B \to \infty$ we get

$$\begin{aligned} t_c &= \frac{1}{|c_{12}|^2}, \\ t_1 &= \frac{1}{|c_1|^2}, \\ t_2 &= \frac{1}{|c_2|^2}, \\ t_{12} &= \frac{2}{|c_1|^2+|c_2|^2}, \\ \tilde{t}_1 &= \frac{1-t_c|c_1|^2}{|c_1|^2+|c_2|^2}+t_c, \\ \tilde{t}_2 &= \frac{1-t_c|c_2|^2}{|c_1|^2+|c_2|^2}+t_c, \end{aligned} \tag{5.35}$$

and the achievable $E_{b,\min}$ (in dB) is

$$E_{b,\min} = 10\log_{10}(t\ln 2), \tag{5.36}$$

with $t = \max\{\min\{t_1, \tilde{t}_1\}, \min\{t_2, \tilde{t}_2\}, t_{12}\}$.

For the wideband slope, we need to find $\mathcal{S}_0(\mathbf{c})$. If $|c_{12}| < \min\{|c_1|, |c_2|\}$, the two users do not cooperate, and in that case we have

$$\mathcal{S}_0(\mathbf{c}) = 2 \qquad \text{CDMA}, \tag{5.37}$$

$$\mathcal{S}_0(\mathbf{c}) = 1 \qquad \text{FDMA}. \tag{5.38}$$

On the other hand, if $|c_{12}| > \min\{|c_1|, |c_2|\}$ the users cooperate. As before, we assume that $|c_1| \leq |c_2|$. If $t_{12} \geq \tilde{t}_1$ in (5.35), the sum rate constraint determines $E_{b,\min}$ for small $E_b - E_{b,\min}$, and by continuity also the wideband slope. Therefore

$$\mathcal{S}_0(\mathbf{c}) = 1 \qquad \text{CDMA}, \tag{5.39}$$

$$\mathcal{S}_0(\mathbf{c}) = \frac{\left(|c_1|^2 + |c_2|^2\right)^2}{2\left(|c_1|^4 + |c_2|^4\right)} \qquad \text{FDMA}, \tag{5.40}$$

where for FDMA we have substituted $N = 2$ in (5.21). Finally, if $t_{12} < \tilde{t}_1$ in (5.35), the spectral efficiency is determined by $\tilde{t}_1^{-1}$ in (5.31) or (5.34) for small $E_b - E_{b,\min}$. Using $R = \tilde{t}_1^{-1}$ in Theorem 9 of [90] and Mathematica for differentiation and limit operations, we arrive at

$$\mathcal{S}_0(\mathbf{c}) = \frac{2|c_{12}|^2(|c_1|^2 + |c_2|^2)}{|c_1|^2(|c_{12}|^2 - |c_2|^2) + 2|c_{12}|^2|c_2|^2} \qquad \text{CDMA},$$

$$\mathcal{S}_0(\mathbf{c}) = \frac{|c_{12}|^2(|c_1|^2 + |c_2|^2)^2}{|c_1|^2(|c_{12}|^2 - |c_2|^2)|c_2|^2 + 2|c_{12}|^2|c_2|^4 + |c_1|^4(|c_{12}|^2 + |c_2|^2)} \qquad \text{FDMA}.$$

Finally we need

$$\begin{aligned}\nabla E_{b,\min}(\mathbf{c}) &= \left(\frac{\partial E_{b,\min}(\mathbf{c})}{\partial |c_1|^2}, \frac{\partial E_{b,\min}(\mathbf{c})}{\partial |c_2|^2}, \frac{\partial E_{b,\min}(\mathbf{c})}{\partial |c_{12}|^2}\right) \\ &= \frac{10}{\ln 10}\left(-\frac{1}{(|c_1|^2 + |c_2|^2)}, \frac{|c_1|^2 - |c_{12}|^2}{(|c_1|^2 + |c_2|^2)(|c_{12}|^2 + |c_2|^2)}, \right. \\ &\qquad \left. -\frac{|c_2|^2}{|c_{12}|^2(|c_{12}|^2 + |c_2|^2)}\right).\end{aligned}$$

The outage wideband slope can now be calculated from (5.13). The result is shown in Fig. 39, which shows that for the two-user case, the difference between the achievable wideband slope of CDMA and FDMA is very small, thus indicating that in the wideband regime, the performance of FDMA based cooperative MAC is almost the same as that of CDMA. This can also be seen from the spectral efficiency versus E_b curve of Fig. 40 where FDMA performs very close to CDMA at low spectral

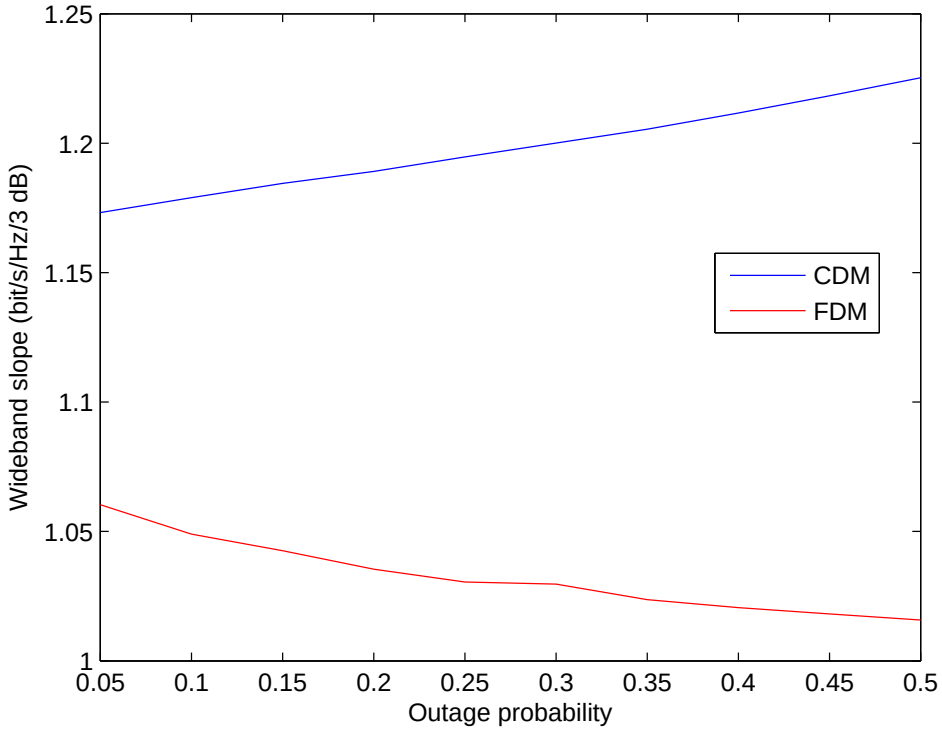


Fig. 39. The wideband slope for the two-user case based on the achievable rate with multiplexed rateless cooperation.

efficiencies. These results, generalized to the N-user case, will be discussed in more details in Section F.

3. The N-user case

In this section, we generalize the bucket-filling interpretation of the two-user case to N users. Define (all quantities depend on $\mathbf{c}$ and time t, but we do not make it explicit for notational convenience)

$\mathcal{D}_j$	:	Set of users that node j has decoded; $\mathcal{D}_j = \{j\}$ for $t = 0$.
$\mathcal{S}$	:	Decoding set. A set of users to be decoded jointly.
$\mathbf{F}_j(\mathcal{U}, \mathcal{S}) \in \mathbb{R}^+$	:	Filled volume at node j for the bucket corresponding to the sum-rate constraint for the users in $\mathcal{U}$, $\mathcal{U} \subseteq \mathcal{S}$.
$r_j(\mathcal{U}, \mathcal{S}) \in \mathbb{R}^+$	:	Fill rate at node j for the bucket corresponding to the sum-rate constraint for the users in $\mathcal{U}$, $\mathcal{U} \subseteq \mathcal{S}$.

In addition, the fill rates $r_j(\mathcal{U}, \mathcal{S})$ are also functions of the current decoding state of users. Expressions for these fill rates for different scenarios (FDMA, CDMA and the limiting case SNR $\rightarrow 0$, both with multiplexed and superposition coding) are provided in Appendix E. At any given time t, user j can decode a set of users $\mathcal{S}$ if the information buckets at user j corresponding to all sum-rate constraints for users $\mathcal{U} \subseteq \mathcal{S}$ are full. In other words, user j can decode a set of users $\mathcal{S}$ if $\mathbf{F}_j(\mathcal{U}, \mathcal{S}) \geq 1 \; \forall \; \mathcal{U} \subseteq \mathcal{S}$. An additional requirement for multiplexed coding is that the decoding state of users should be such which allow the joint decoding of $\mathcal{S}$. For example, for $N = 3$, $\mathcal{S} = \{1\}$ cannot be decoded at user 3 if $\mathcal{D}_1 = \mathcal{D}_2 = \{1, 2\}$, and $\mathcal{D}_3 = \{3\}$. In general, a decoding set $\mathcal{S}$ is valid if there exists a subset $\mathcal{V}$ of users such that[4]

$$\mathcal{S} \;=\; \bigcup_{j \in \mathcal{V}} \mathcal{D}_j(t). \tag{5.41}$$

Finally, successful decoding occurs at the collector after time $T(\mathbf{c})$, which is the minimum time it takes to obtain $\mathcal{D}_{N+1} = \{1, \ldots, N+1\}$. The time $T(\mathbf{c})$ can be obtained by the following algorithm:

1: Initialize: $T(\mathbf{c})$=0, $\mathcal{D}_j = \{j\}$, and $\mathbf{F}_j(\mathcal{U}, \mathcal{S}) = 0$, $\forall \; \mathcal{U} \subseteq \mathcal{S}$, $\mathcal{S} \subseteq \{1, \ldots, N\}$,

[4]For superposition coding, all decoding sets are valid.

$j = 1, \ldots, N+1$.

2: **While** $|\mathcal{D}_{N+1}| < N$

3: Update fill rates $r_j(\mathcal{U}, \mathcal{S})\ \forall\ \mathcal{U} \subseteq \mathcal{S},\ \mathcal{S} \subseteq \{1, \ldots, N\},\ j = 1, \ldots, N+1$ – See Appendix E.

4: Calculate decoding times: $\tau_j(\mathcal{S}) = \max_{\mathcal{U} \subseteq \mathcal{S}} \frac{1 - \mathbf{F}_j(\mathcal{U}, \mathcal{S})}{r_j(\mathcal{U}, \mathcal{S})},\ \forall\ S \subseteq \{1, \ldots, N\},\ j = 1, \ldots, N+1$.

5: Calculate minimum fill time: $\tau = \min \left\{ \tau_j(\mathcal{S}) \;\middle|\; \tau_j(\mathcal{S}) \geq 0 \text{ and } \mathcal{S} \text{ is valid} \right\}$.

6: Update volumes: $\mathbf{F}_j(\mathcal{U}, \mathcal{S}) = \mathbf{F}_j(\mathcal{U}, \mathcal{S}) + \tau r_j(\mathcal{U}, \mathcal{S})$, for all $\mathcal{U}, \mathcal{S}$, and j.

7: Update decoding sets: $\mathcal{D}_j = \mathcal{D}_j \bigcup \left\{ \mathcal{S} \;\middle|\; \tau = t_j(\mathcal{S}) \right\}, j = 1, \ldots, N+1$.

8: Update time: $T(\mathbf{c}) = T(\mathbf{c}) + \tau$.

9: **end while**

Then for the given channel coefficients $\mathbf{c}$, the achievable spectral efficiency is equal to $\frac{1}{T(\mathbf{c})}$. In addition, if (for either FDMA, CDMA or SNR $\to 0$, and superposition or multiplexed coding)

$$R(p) \;\; = \;\; \max \left\{ R \left| \Pr \left(R > \frac{1}{T(\mathbf{c})} \right) \leq p \right. \right\}, \tag{5.42}$$

the spectral efficiency $R(p)$ is achievable with outage probability p using rateless coding. While this may seem obvious, one does need a formal proof because of the compound nature of the channel – the interested reader is referred to [91]. Let $T_0(\mathbf{c})$ be $T(\mathbf{c})$ for SNR $\to 0$. Then the outage minimum energy per bit achievable with rateless coding can be found as

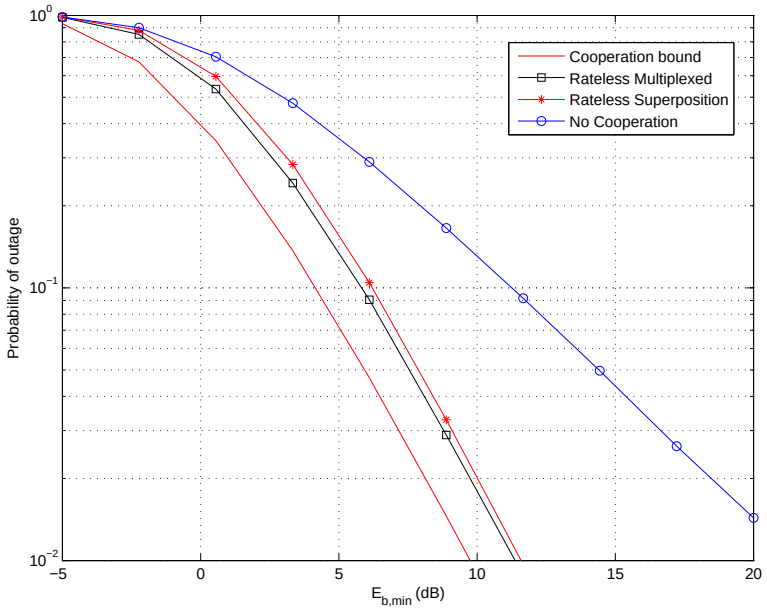
$$E_{b,\min}(p) \;\; = -1.59 + 10 \log_{10} \left(\max \left\{ T \;\middle|\; \Pr \left(T_0(\mathbf{c}) > T \right) \leq p \right\} \right). \tag{5.43}$$

F. Numerical Results for Capacity

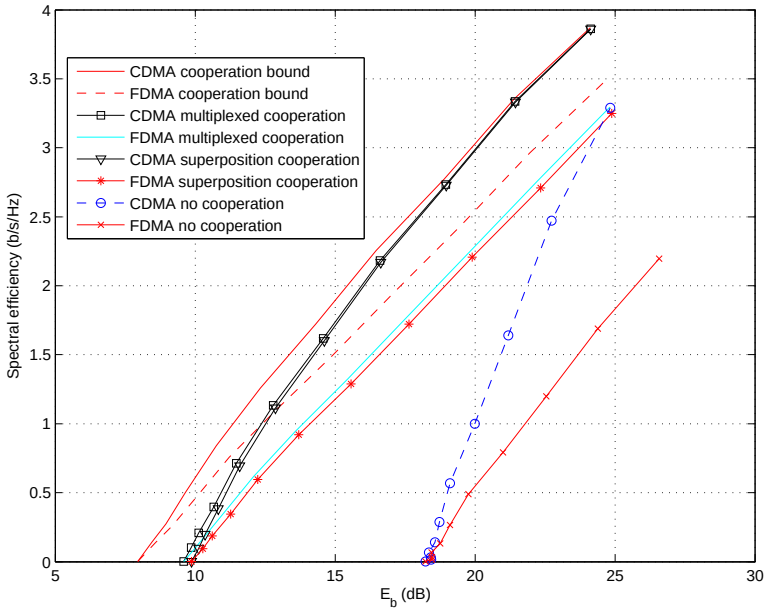
In this section, we provide numerical results for outage capacity when all channel gains experience i.i.d. Rayleigh fading with unit variance. We plot the outage capacity versus $E_{b,\min}$, as well as the outage spectral efficiency R versus E_b for $N = 2, 4, 8$ users in Figs. 40–42. Some observations that can be made are listed below.

- **Cooperation Gain:** There is a very large gain in $E_{b,\min}$ from cooperation, in the order of tens of dB. In fact, when the number of users increase, the required $E_{b,\min}$ for a particular probability of outage increases without cooperation, but *decreases* with cooperation. However, the gain from cooperation quickly disappears as the spectral efficiency increases, at least for the CDMA case. From Fig. 42, it can be seen that at a spectral efficiency of 1.5 bits/s/Hz, there is almost no gain from cooperation. This can be explained by the fact that CDMA without cooperation has a wideband slope of 2 while the wideband slope of any cooperative scheme is much smaller (see Fig. 36). In fact, from (5.15), (5.20) and Fig. 36, it can be expected that as the number of users increase further, the gain in $E_{b,\min}$ will increase, but the spectral efficiency needed to achieve that gain will become smaller and smaller.

- **FDMA versus CDMA**: As stated before, FDMA and CDMA have the same $E_{b,\min}$. On the other hand, the wideband slope for CDMA is larger than that for FDMA, but only slightly, as indicated by Figs. 36 and 39. This is also illustrated in Figs. 40–42, where it can be observed that if the spectral efficiency is kept very low, there is no loss from using FDMA with cooperation. However, even at higher spectral efficiencies the loss is limited to a only a few dBs. This is starkly different from the no cooperation case, where FDMA operation performs much worse than CDMA. For FDMA, cooperation also pays off at higher spectral

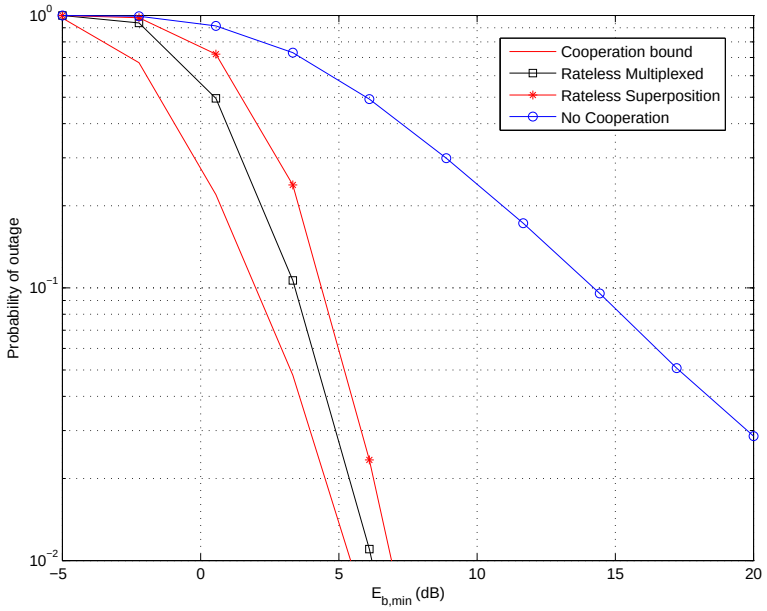


(a)

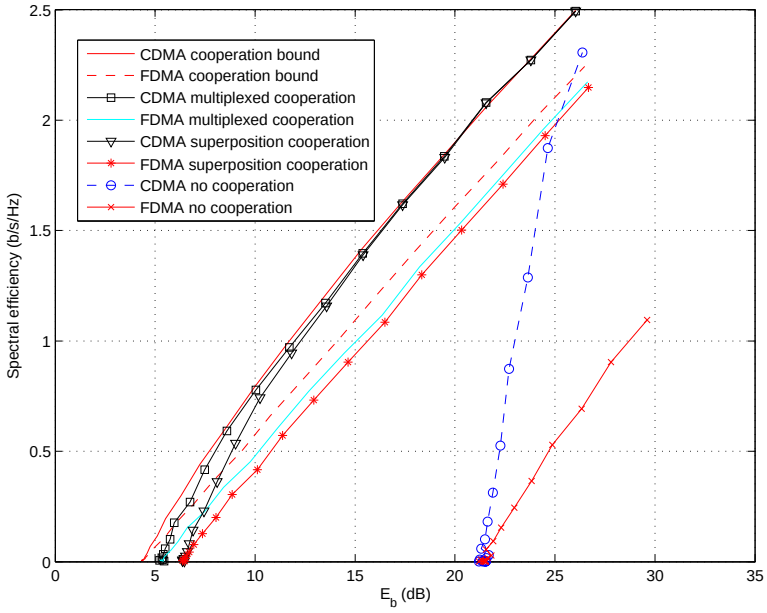


(b)

Fig. 40. Outage performance in two-user MAC under i.i.d. Rayleigh fading (a) Outage probability versus $E_{b,\min}(p)$ and (b) Outage Spectral efficiency versus E_b for an outage probability of 0.02.

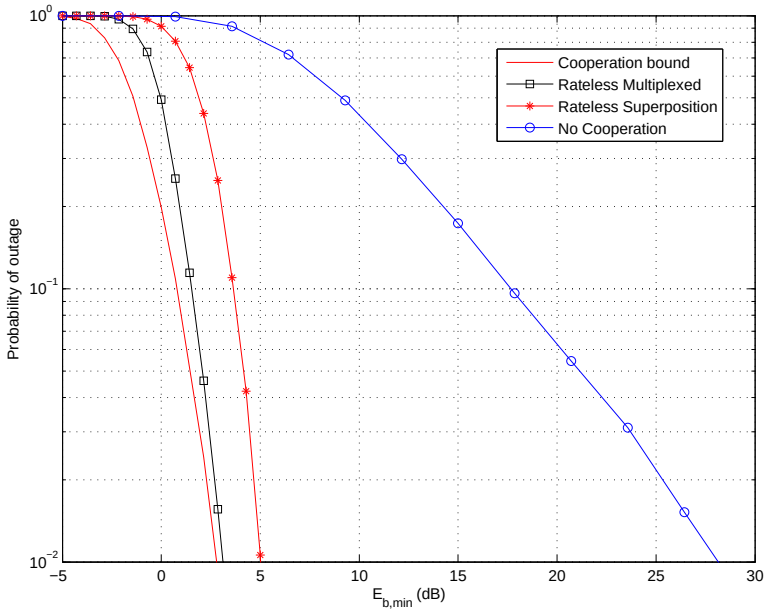


(a)

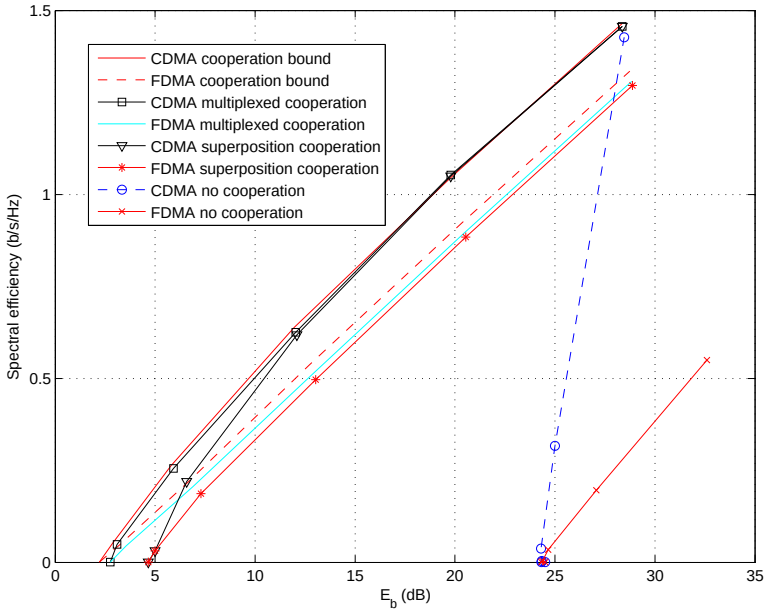


(b)

Fig. 41. Outage performance in four-user MAC under i.i.d. Rayleigh fading (a) Outage probability versus $E_{b,\min}(p)$ and (b) Outage Spectral efficiency versus E_b for an outage probability of 0.02.



(a)



(b)

Fig. 42. Outage performance in eight-user MAC under i.i.d. Rayleigh fading (a) Outage probability versus $E_{b,\min}(p)$ and (b) Outage Spectral efficiency versus E_b for an outage probability of 0.02.

efficiencies, as opposed to CDMA.

- **Multiplexing versus Superposition**: For eight users the gain in $E_{b,\min}$ from using multiplexing over superposition is around 2.5 dB. Furthermore, for low outage probability, multiplexed coding comes within 0.1 dB of the outer bound (i.e., within 0.1dB of capacity). The gain in $E_{b,\min}$ increases with the number of users, so there might be an even larger gain with more users. On the other hand, Fig. 42 shows that the spectral efficiency has to be low to realize the gain from multiplexing.

G. Practical Multiplexed Rateless Cooperation

In this section, we describe how a rateless coded cooperative strategy for a half-duplex MAC can be implemented using multiplexed Raptor codes [79]. As a first step towards developing practical coding strategies for a cooperative MAC, we present a linear programming approach towards designing multiplexed Raptor codes. We then briefly explain how the designed codes are used in implementing the cooperative coding scheme of Section E-3 and present simulation results.

1. Multiplexed Raptor code design

A design methodology for Raptor codes specifically geared towards arbitrary binary input symmetric channels is provided in [80]. However, it is shown that Raptor codes on AWGN channels are not universal, i.e., the capacity approaching code is a function of the channel signal-to-noise ratio (SNR). Since in our setup we assume that the transmitters have no knowledge of the channel fading coefficients, it is not clear what channel gains the multiplexed Raptor code should be designed for. We therefore simplify the design process by assuming transmission over a binary erasure channel.

The motivation for this simplification is the fact that Raptor codes designed for the binary erasure channel work reasonably well on the AWGN channel [80, 102]. The optimized codes are then used to simulate practical rateless MAC cooperation with all channels having AWGN. The scheme is found to operate close to the theoretical limit as shown in Section G-3.

Encoding in Raptor codes is achieved by first precoding the original message with a high-rate LDPC code to obtain a k-bit intermediate message. In the following, we will refer to these intermediate message bits as the input bits. The intermediate message is then encoded using an LT code [82]. For LT encoding an output bit, first a degree d is chosen according to the distribution defined by the polynomial $\mathbf{\Omega}(x) = \sum_{d=1}^{D} \Omega_d x^d$, where Ω_d is the probability of the degree being d and D is the maximum node degree. Then d random input bits are chosen and added modulo 2 to obtain the output bit.

Consider first single user transmission, where the k-bit intermediate message is LT encoded to form an n-bit codeword and transmitted over a binary erasure channel with capacity C. If n is large, the decoder receives nC non-erasures, which it uses to recover the intermediate message. One can define an equivalent LT decoding graph with k input (variable) nodes and nC output (check) nodes, with the output node degree distribution given by $\mathbf{\Omega}(x)$. Since the neighbors of an output node are chosen randomly, it induces a degree distribution on the input nodes which can be evaluated as [80]

$$\mathbf{I}(x) \approx \exp\left(\frac{\mathbf{\Omega}'(1)(x-1)}{r}\right), \tag{5.44}$$

where we define $r = k/nC$ as the inverse overhead with $0 \leq r \leq 1$ and $\mathbf{\Omega}'(1)$ is the first derivative of $\mathbf{\Omega}(x)$ evaluated at 1. The approximation in (5.44) becomes exact as k approaches infinity. Decoding of the intermediate message is accomplished by

running the iterative message-passing algorithm on the LT decoding graph. According to [73, 103], if x is the probability that a message from an input node to the output node is an erasure, the condition required to recover a fraction $1-\delta$ of the intermediate message bits[5] for $k \to \infty$ is given by $\mathbf{I}'\left(1 - \frac{\mathbf{\Omega}'(1-x)}{\mathbf{\Omega}'(1)}\right) < \mathbf{I}'(1)x$ for all $x \in [\delta, 1]$. Using (5.44), this condition can be simplified as

$$\mathbf{\Omega}'(1-x) + r\ln(x) > 0, \qquad \forall \quad x \in [\delta, 1]. \tag{5.45}$$

Note that in order to recover the message in as few transmissions as possible, one should minimize the number of output bits required to recover the message, which is equivalent to maximizing r. However, the degree distribution obtained by solving the optimization problem performs poorly in practice for finite k, especially since the solution results in the fraction of degree one nodes approaching zero – in the absence of degree one nodes, the message passing algorithm does not work . Instead of (5.45), we use the modified heuristic convergence condition for finite k presented in [79] which is given by

$$\mathbf{\Omega}'(1-x) + r\ln\left(x - \sqrt{\frac{x}{k}}\right) > 0, \qquad \forall \quad x \in [\delta, 1]. \tag{5.46}$$

We now discuss multiplexed codebook design and consider the case when N users need to be encoded. We assume that the message of each user is encoded by an independent LDPC code, thus obtaining N length-k intermediate messages. The overall intermediate message of length-Nk is then encoded by an LT code characterized by the degree distribution polynomial $\mathbf{\Omega}_N(x)$. The objective of the design process is to choose the degree distribution so that 1) the overall length-Nk message is recoverable close to the channel capacity, and 2) if the decoder were to know messages corre-

[5]It is assumed that the fixed rate LDPC precode is capable of recovering the remaining fraction δ of the intermediate bits.

sponding to an arbitrary subset of users beforehand, the remaining sets of messages is also recoverable as close to channel capacity as possible. Consider that the decoder has, through some other means, complete knowledge of the message of $N-m$ users, $m = 1, \ldots, N$. The decoder needs to recover the remaining m sets of messages. In iterative BP decoding, the edges emanating from the already available $N-m$ messages do not play any other role than sign reversals at the check nodes, and can therefore be removed from the decoding graph. The remaining edges induce a degree distribution $\mathbf{\Omega^m}(\mathbf{x})$ which is related to $\mathbf{\Omega}_N(x)$ as follows. For a degree-d output node, the probability that it has d_m edges connected to the unknown subset of messages is given by

$$p(d, d_m) = \binom{d}{d_m} \left(\frac{m}{N}\right)^{d_m} \left(1 - \frac{m}{N}\right)^{d-d_m} u[d - d_m],$$

with $d_m = 0, 1, \ldots, D$, $u[k] = 1$ when $k \geq 0$, and $u[k] = 0$ otherwise. Consequently, when the edges from the known set of messages are removed, the fraction of output nodes with degree d_m is given by $\Omega_{m,d_m} = \sum_{d=1}^{D} \Omega_{N,d} p(d, d_m)$. The induced degree distribution polynomial $\mathbf{\Omega}^m(x) = \sum_{d_m} \Omega_{m,d_m} x^{d_m}$ for the m unknown messages can be computed as

$$\begin{aligned}
\mathbf{\Omega}^m(x) &= \sum_{d=1}^{D} \Omega_{N,d} \left(1 - \frac{m}{N}\right)^d \sum_{d_m=0}^{d} \binom{d}{d_m} \left(\frac{m}{N-m} x\right)^{d_m} \\
&= \sum_{d=1}^{D} \Omega_{N,d} \left(1 - \frac{m}{N}\right)^d \left(1 + \frac{m}{N-m} x\right)^d \\
&= \mathbf{\Omega}_N \left(\frac{N - m + mx}{N}\right).
\end{aligned}$$

Let r_m be the inverse overhead associated with recovering m sets of messages when $N-m$ messages are already known. Then the design requirement for recovering a fraction $1-\delta$ of the mk intermediate bits when $(N-m)k$ intermediate bits are already

known can be obtained from (5.46) and is given by

$$\frac{m}{N}\mathbf{\Omega}'_N\left(\frac{N-mx}{N}\right)+r_m\ln\left(x-\sqrt{\frac{x}{mk}}\right)>0,\quad\forall\quad x\in[\delta,1]. \tag{5.47}$$

For the multiplexed code design, the design requirement in (5.47) should be satisfied for all $m=1,\ldots,N$. As described above, for $N=1$, the design goal is to find the maximum r_1 for which there exists a valid degree distribution satisfying (5.47). For $N>1$, it is hard to define an objective function for the best frame error rate performance. Instead, we consider a possibly sub-optimum objective function $\sum_{m=1}^{N} r_m$ which is linear in r_m, and is found to result in a frame error rate close to the theoretical limit. By requiring the inequality (5.47) to hold for x belonging to discretized points in the interval $[\delta,1]$, one obtains a sequence of linear inequalities. In addition, we have the trivial constraints

$$\mathbf{\Omega}_N(1)=1, \tag{5.48}$$

$$\Omega_{N,d}\geq 0,\quad d=1,\ldots,D, \tag{5.49}$$

$$0\leq r_m\leq 1,\quad m=1,\ldots,N, \tag{5.50}$$

all of which are linear in terms of $\Omega_{N,d}$ and r_m. Thus the optimization problem can be solved using linear programming. Table VII shows the r_m's obtained by solving the linear programming problem for various N when the maximum output node degree D is limited to 100. It is seen that the resulting Raptor codes are not optimally multiplexed, since when N is large, r_m for $m<N$ is much smaller than one. However, as indicated by our simulations in Section G-3, rateless multiplexed cooperation using the degree distributions corresponding to the parameters in Table VII perform well in practice with a gap of 0.52 dB and 1.1 dB to the theoretical limit for the two- and four-user case, respectively.

Table VII. The inverse overheads r_m for several N's and fixed $\delta = 0.01$.

N	r_1	r_2	r_3	r_4
1	0.985	–	–	–
2	0.699	0.925	–	–
3	0.539	0.817	0.904	–
4	0.482	0.717	0.821	0.871

2. Coding scheme

As in Section E, we assume that all users transmit at the same rate of R bits per complex sample, and have the same power constraint P. Since we implement the FDMA based half-duplex cooperation, different users are allocated to non-overlapping frequency bands, the width of each being $1/N$ of the overall transmission bandwidth. To obtain an overall transmission rate of R, each user must transmit at a rate NR with power constraint NP over its allocated frequency band. A user first precodes its message with an LDPC code of fixed rate R_L, LT encodes the length-k intermediate message, QPSK modulates the resulting bits using the constellation points $\left\{\pm\sqrt{\frac{NP}{2}}, \pm\sqrt{\frac{NP}{2}}\right\}$ and then transmits to the collector. At the same time, it overhears the transmission of the remaining users. Since the frequency bands allocated to the users are non-overlapping, the transmission it receives from other users do not interfere with each other, and hence it does not need to employ multi-user detection. It is assumed that all users, in addition to the collector have complete knowledge of the encoding graphs. When a user has received the minimum number of symbols required to decode another user, it forms an equivalent decoding graph and performs iterative belief propagation decoding. If a user at any given time has recovered m messages including its own, it uses the degree distribution $\mathbf{\Omega}_m(x)$, obtained by the

linear programming approach presented in Section G-1, to LT encode the intermediate messages. We assume that a user knows whether it has decoded another message correctly, something which can be achieved in practice by employing the likes of CRC codes. If decoding fails, a user waits to receive more symbols before attempting decoding again. Ideally, it should make a new decoding attempt every time it receives a new symbol. However in order to avoid the computation cost associated with making so many decoding attempts, we let the users receive an additional 2.5% (of the information theoretic minimum) symbols before starting a new decoding process.

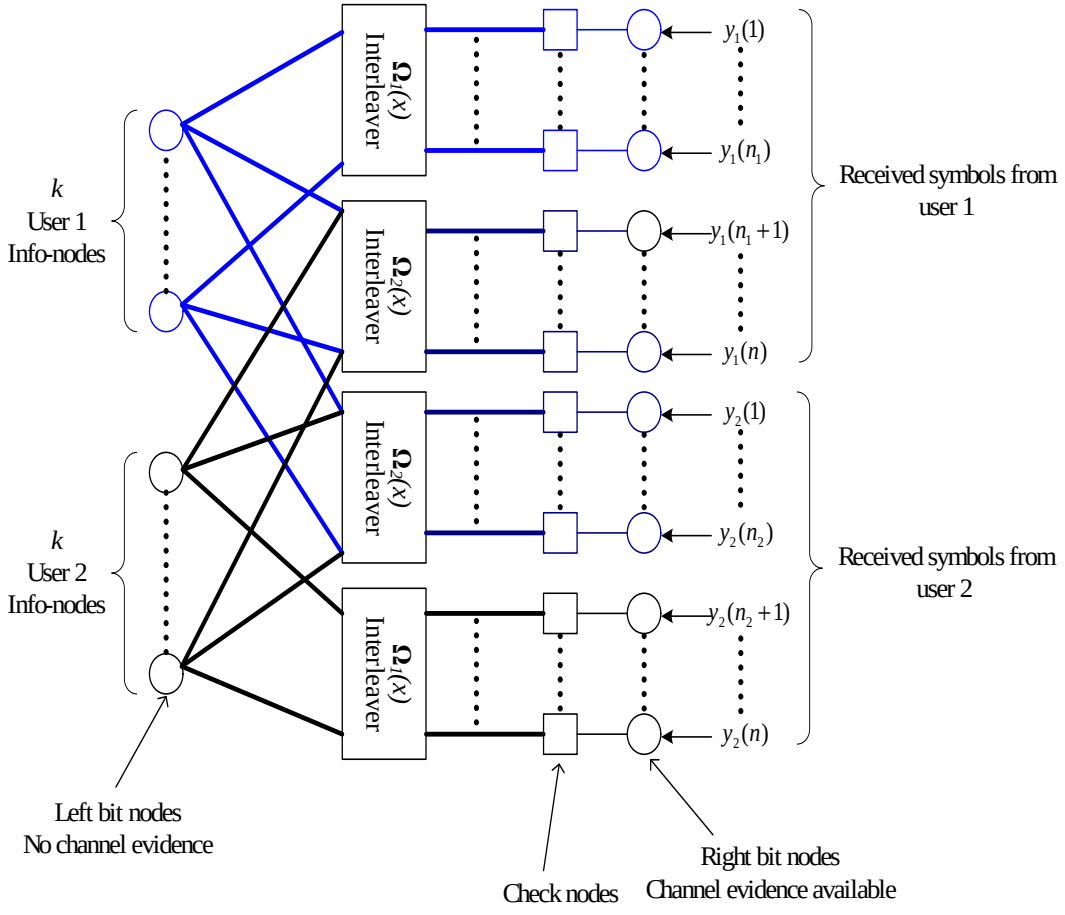


Fig. 43. Decoding graph at the collector for the two-user case when user i, $i = 1, 2$, decodes the other user after n_i symbols. The portion of the graph corresponding to the LDPC pre-codes is not shown for clarity.

Decoding at the collector is very similar to that at the users, except that it makes a single decoding attempt after it has received the entire block of length $n = \frac{kR_L}{NR}$ symbols per user. The equivalent decoding graph at the collector for $N = 2$ when user i, $i = 1, 2$, decodes the other user after n_i symbols is shown in Fig. 43. Note that information theoretically the two decoding times should be equal because of channel reciprocity. However, in practice, the two users might decode each other at different times, i.e., in general $n_1 \neq n_2$. For the purpose of clarity, the figure does not show the portion of the decoding graph corresponding to the LDPC pre-codes. The equivalent decoding graphs at the collector as well as the at the users for $N > 2$ can be constructed in a similar manner as in Fig. 43.

3. Simulation results

We present simulation results for the practical rateless cooperation scheme for the two- and four-user cases for a fixed transmission rate of $R = 0.05$ bits per sample. As a performance measure, we consider the frame error rate when all channels experience independent slow Rayleigh flat fading. All messages are of length 9500 bits and are precoded using a rate-$\frac{95}{100}$ LDPC code to obtain length-$k = 10,000$ intermediate messages. The maximum number of decoding iterations at the users are limited to 100, and at the collector to 200. As in the theoretical outage analysis, a frame error is declared when the collector is unable to decode at least one of the users' messages. If for given channel conditions $\mathbf{c}$ and a given E_b, $R_a(E_b, \mathbf{c})$ is the information theoretic achievable rate calculated by the bucket filling interpretation of Section E-3, the frame error rate for the practical scheme can be evaluated as

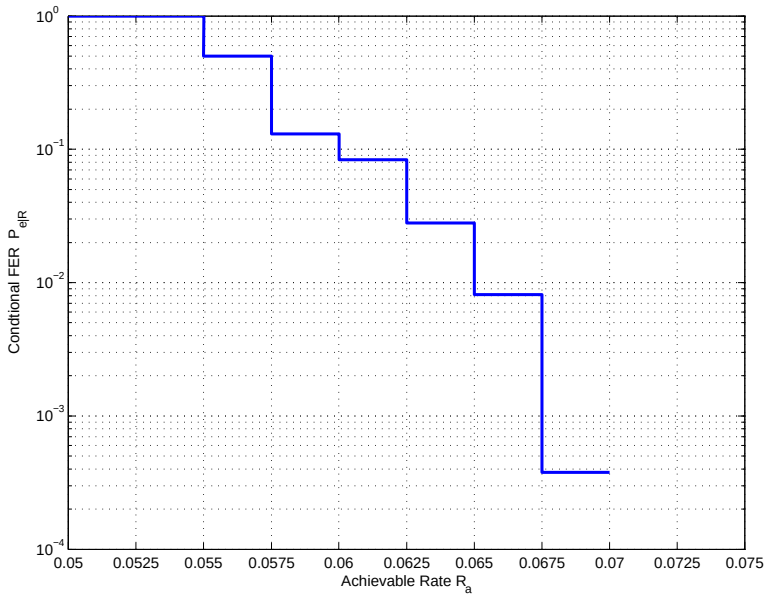
$$P_e(E_b) = \sum_{i=0}^{\infty} P_{e|R}(i\triangle, (i+1)\triangle) \Pr(i\triangle \leq R_a(E_b, \mathbf{c}) < (i+1)\triangle), \tag{5.51}$$

where $\triangle > 0$, and $P_{e|R}(i\triangle, (i+1)\triangle)$ is the conditional probability of frame error for the practical coding scheme given the information theoretic achievable rate lies in the interval $[i\triangle, (i+1)\triangle)$. Note that we have assumed that this conditional probability of frame error is independent of E_b which is true if $\triangle$ is small. We obtain $P_{e|R}$ through Monte-Carlo simulations of the practical multiplexed rateless cooperation scheme. Fig. 44 (a) shows this conditional probability of frame error for the two-user case for the interval $R_a \in [0.05, 0.07)$ with $\triangle = 0.0025$. In order to calculate the overall probability of frame error in (5.51), we assume that $P_{e|R}$ for $R_a \geq 0.07$ reaches an error floor of 3.8×10^{-4} (corresponding to the frame error rate of the interval $R_a \in [0.0675, 0.07)$) and that the frame error rate for $R_a < 0.05$ is (obviously) one. The probability of frame error for the two-user case when all channels experience slow Rayleigh fading is shown in Fig. 44 (b). At a frame error rate of 2×10^{-2}, the simulation results are observed to lose 0.52 dB from the theoretical FDMA cooperation limit, and are 7.75 dB better than the no cooperation bound (with CDMA).

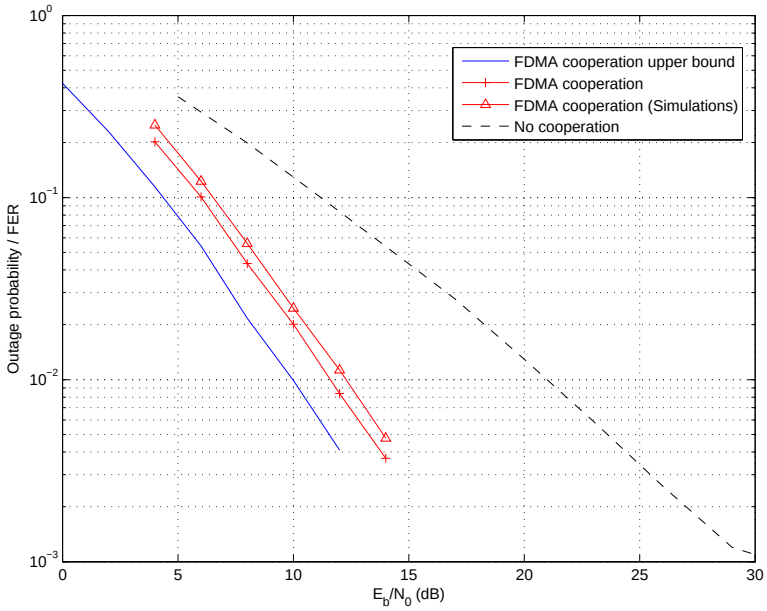
Fig. 45 shows the same results for the four-user case, which indicate that the scheme with multiplexed Raptor codes loses only 1.1 dB and 14.66 dB from the theoretical FDMA cooperation limit and the no cooperation bound, respectively.

H. Summary

The results in this chapter can be used as indicators of how to design energy efficient networks, in particular when the channels are unknown or uncertain. Foremost, the network should use cooperative diversity; the energy gain from this is huge. However, to realize this gain, the network has to operate in the (very) low power regime, either through slow transmission or by using ultra-wideband transmission. Furthermore, the nodes do not need to be full-duplex capable to realize the energy saving. Finally, by

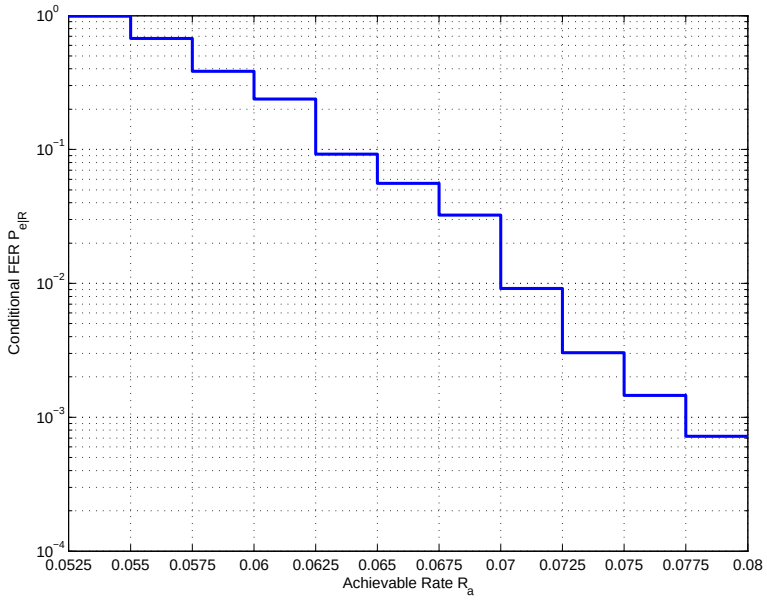


(a)

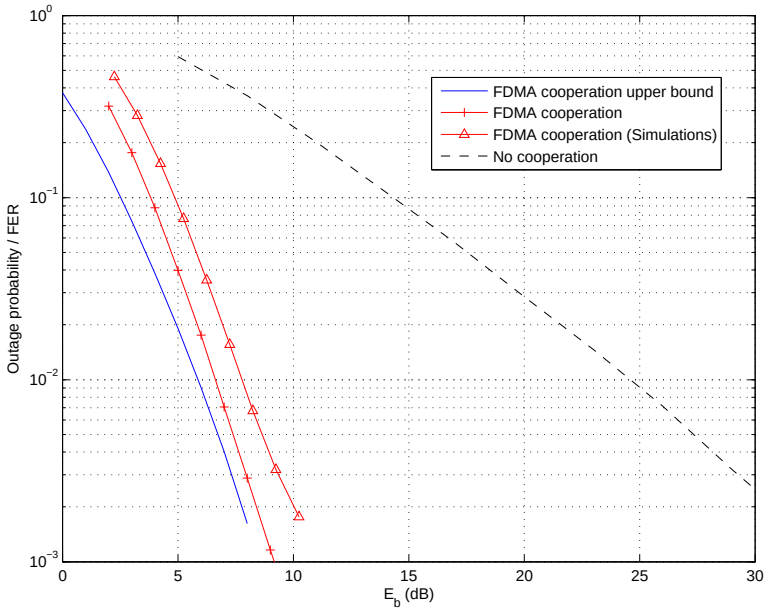


(b)

Fig. 44. Simulation results for two-user case at fixed transmission rate of $R = 0.05$ b/s. (a) Conditional probability of frame error $P_{e|R}$ versus the achievable rate with $\triangle = 0.0025$, (b) Probability of frame error versus E_b compared to theoretical bounds.



(a)



(b)

Fig. 45. Simulation results for four-user case at fixed transmission rate of $R = 0.05$ b/s. (a) Conditional probability of frame error $P_{e|R}$ versus the achievable rate with $\triangle = 0.0025$, (b) Probability of frame error versus $E_b$compared to theoretical bounds.

using multiplexed coding, an additional gain of 2-3 dB can be achieved, with energy consumption close to the theoretical limit. Extending on the theoretical results, we have presented a practical scheme for the cooperative MAC which utilizes multiplexed Raptor codes. Simulation results indicate a loss of 0.52 dB and 1.1 dB from the theoretical limit for the two- and four-user case, respectively.

We have only included numerical results for up to eight nodes, and simulation results for up to four nodes. The reason for this is certainly complexity, but also that the model used might not be realistic for very large networks. If the network is large, nodes are probably spread out over a large area, and path loss is then relevant, and should be taken into account. In such a situation, a combination of cooperative diversity (e.g., multiplexed codes) and routing seems to be the right approach. This could be a direction for further research.

CHAPTER VI

THE COGNITIVE RADIO CHANNEL

A. Introduction

With the growing trend in the use of wireless systems, it has been observed that the current licensed spectrum is severely under-utilized. In order to use the licensed spectrum more efficiently, the idea of deploying secondary wireless devices has been proposed – with the secondary devices using the same frequency band as the existing primary systems. In order to have a minimal effect on the operation and performance of the primary systems, these secondary devices have to be opportunistic in nature. In other words, they should be able to "cognitively" adapt to their environment to utilize channel resources when they become available, while at the same time they should be able to communicate effectively with their respective base stations. Because of this required cognition capability, the term "cognitive radio" has been widely used in the literature for these secondary devices.

The simplest form of cognitive radio channel (CRC) consisting of one primary and one secondary/cognitive user is shown in Fig. 46. Each user wishes to communicate some information to its respective base-station. However, since the two users share the same frequency band, the signal transmitted by a user interferes with the transmission of the other. A number of works have focused on deriving the information theoretic achievable rates for the CRC when the primary user message is known at the cognitive user non-causally [104, 105, 106], as well as causally [107]. For the non-causal case, it was shown in [104] that when the channel from the cognitive user to its respective base-station is stronger than the cognitive user to primary receiver channel (referred to as the low-interference regime), a scheme which achieves capacity involves dirty-

paper coding (DPC) [31]. For the general CRC as well, the inner bounds on the capacity region rely heavily on DPC. Almost all achievable schemes in the literature rely on the following methodology: Since it is assumed that the cognitive user knows the primary user message before it begins transmitting, it can dirty-paper code its own message by treating a scaled version of the primary user transmission as the side-information. As a result, the secondary receiver sees no (or partial) interference from the primary user. At the same time, the cognitive user allocates some of its power to transmitting the primary user message in order to reduce the effect of interference caused by its own transmission. In this chapter, we design and simulate a DPC based coding scheme for the cognitive user channel which relies on the methodology described above. As a first step towards developing a practical coding strategy, we assume that the cognitive user has knowledge of the primary user message as well as its codebook before the transmissions begin. In addition, we consider the low interference regime where the channel gain from the cognitive user to its base station is stronger than it is to the primary base station. This scenario is of practical interest since the cognitive user will typically be closer to its base station than the primary base station. As mentioned earlier, the cognitive radio should be minimally intrusive in the operation of the primary system. We consider the extreme situation where the introduction of the cognitive radio should have no effect on the primary user's operation and performance whatsoever, i.e. the primary system should operate as if there was no cognitive user in the system. For the primary user, we use a low density parity-check (LDPC) code with the codeword bits mapped to a 4-PAM constellation, whereas for the cognitive user, we propose a multi-level DPC coding scheme which uses an LDPC code as the channel coding, and trellis coded quantization (TCQ) code as the source coding component. Simulations indicate that at a transmission rate of 1.0 bits/sample (b/s) and a block length of 50,000, the proposed DPC scheme

performs within 0.78 dB of the theoretical limit.

The chapter is organized as follows. In Section B, we present the model for the CRC. In Section C, we discuss the DPC based coding scheme and present the corresponding performance limits. Section D forms the major part of this chapter, in which we present the proposed DPC scheme, and discuss the design issues. In Section E, we present the simulation results, and finally provide a summary in Section F.

B. Channel Model

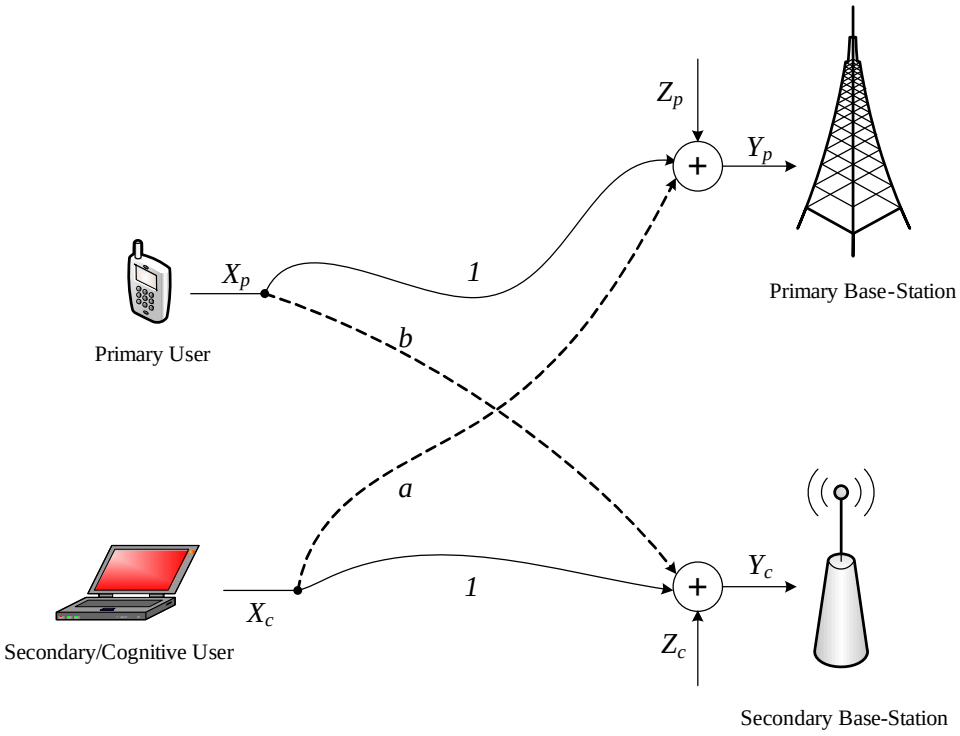


Fig. 46. The cognitive radio channel.

Without loss of generality, we consider the cognitive radio channel in its standard form [104] as depicted in Fig. 46. The signal received at the primary receiver at time

$n = 1, \ldots, L$ (L being the transmission block length) is given by

$$Y_p[n] = X_p[n] + aX_c[n] + Z_p[n],$$

where $X_p[n]$ and $X_c[n]$ are the signals transmitted from the primary and cognitive user, respectively, $Z_p[n]$ is the unit variance additive white Gaussian noise (AWGN), and a is the channel gain from the cognitive user to the primary receiver. In the standard form of the cognitive channel, low interference regime implies $a < 1$. Similarly, the signal received at the secondary base station is given by

$$Y_c[n] = X_c[n] + bX_p[n] + Z_c[n],$$

where $Z_c[n]$ is once again unit variance AWGN and b is the channel gain from the primary user to the secondary base station. The power constraints at the primary and cognitive users are given by

$$\begin{aligned} \frac{1}{L}\sum_{n=1}^{L} X_p^2[n] &\leq P_p \ , \quad \text{and} \\ \frac{1}{L}\sum_{n=1}^{L} X_c^2[n] &\leq P_c, \end{aligned}$$

respectively. For the sake of simplicity, we assume that the source transmissions $X_p[n]$ and $X_c[n]$, as well as the channel coefficients a and b are real. Extension to complex base-band is relatively straightforward. In addition, we assume, as mentioned earlier, that that the cognitive user has non-causal knowledge of the primary user's transmissions, i.e., the cognitive user has perfect knowledge of $X_p[n]$, $n = 1, \ldots, L$, before the transmissions even begin.

C. Coding Scheme and Performance Limits

In this section, we briefly describe a coding scheme for the CRC [104], and provide the information theoretical performance limits. The primary user encodes its length-k message $\mathbf{m}_p$ to a length-N codeword using a rate-$R_p^l = \frac{k}{N}$ LDPC code. The coded bits are then mapped to the symbols $X_p[n]$, $n = 1, \ldots, L$, which belong to a constellation of size $M = 2^m$. Thus the total number of transmitted symbols is $L = \frac{N}{m}$, with the overall transmission rate from the primary user given as $R_p = mR_p^l$ b/s. The average transmission power is given by $P_p = \frac{1}{L}\sum_{n=1}^{L} X_p^2[n]$. The primary base station estimates its received signal-to-noise ratio (SNR), and attempts to decode $\mathbf{m}_p$ using belief-propagation (BP) algorithm on the LDPC decoding graph. *We assume that the primary user always transmits with the minimum power required for the BP decoder to satisfy a given bit-error rate (BER) requirement, with the decoder assuming that the noise (plus any interference) has Gaussian statistics.*

On the other hand, the cognitive user allocates a fraction γ of its power for transmitting the coded symbols of the primary user – it can do so since it is assumed to know the primary user message as well as its codebook. With the remaining power, it encodes its message $\mathbf{m}_c$ using a DPC scheme (the details of which will be discussed in Section D) with

$$S[n] = \left(b + \sqrt{\frac{\gamma P_c}{P_p^*}}\right) X_p[n]$$

treated as the known interference/side-information. The dirty paper coded output $\tilde{X}_c[n]$ satisfies

$$\frac{1}{L}\sum_{n=1}^{L} \tilde{X}_c^2[n] = (1-\gamma)P_c$$

and the cognitive user transmits

$$X_c[n] = \tilde{X}_c[n] + \sqrt{\frac{\gamma P_c}{P_p}} X_p[n].$$

A key property of the DPC encoder is that the coded output $\tilde{X}_c$ is uncorrelated with the side-information [31] and is therefore uncorrelated with $X_p[n]$ – a property verified by the simulations of our DPC scheme of Section D. Thus, the total transmission power constraint $\frac{1}{L}\sum_{n=1}^{L} X_c^2[n] = P_c$ is satisfied. The signals received at the primary and secondary base station are given as

$$\begin{aligned} Y_p[n] &= \left(1 + a\sqrt{\frac{\gamma P_c}{P_p}}\right) X_p[n] + a\tilde{X}_c[n] + Z_p[n], \\ Y_c[n] &= \tilde{X}_c[n] + \left(b + \sqrt{\frac{\gamma P_c}{P_p^*}}\right) X_p[n] + Z_c[n], \end{aligned}$$

respectively. The primary base station is assumed to be oblivious to the presence of the cognitive user, and therefore, it treats the term $a\tilde{X}_c[n]$ as unknown interference. Since the interference term $a\tilde{X}_c[n]$ is uncorrelated with $X_p[n]$, the received SNR at the primary base station is given as

$$SNR_p = \frac{\left(1 + a\sqrt{\frac{\gamma P_c}{P_p}}\right)^2 P_p}{1 + a^2(1-\gamma)P_c}. \tag{6.1}$$

In order for the SNR at the primary base station to remain the same as in the absence of the cognitive user, the power allocation parameter γ should be chosen such that

$$SNR_p = \frac{\left(1 + a\sqrt{\frac{\gamma P_c}{P_p}}\right)^2 P_p}{1 + a^2(1-\gamma)P_c} = P_p$$

which yields the solution [104]

$$\gamma^*(P_p, P_c) = \left(\frac{\sqrt{P_p}\left(\sqrt{1 + a^2 P_c(1+P_p)} - 1\right)}{a\sqrt{P_c}(1+P_p)}\right)^2. \tag{6.2}$$

In short, the cognitive user compensates for the added interference at the primary receiver by allocating a fraction γ^* of its power for transmitting the primary user message. Note that even though the SNR at the primary destination remains the same as that without the cognitive user, the received signal power does not. Thus, when calculating the log-likelihood ratios for LDPC decoding, the primary receiver needs to account for this increased signal strength[1]. At first glance, this might indicate that the primary receiver has to modify its decoding process. However, note that even in the absence of the cognitive user, the primary receiver has to estimate the received signal strength (possibly through some pilot symbols), and hence one can argue that the increased signal strength does not affect the operation of the primary decoder. Additionally, if the BER of the primary decoder were a function of only the SNR, choosing $\gamma = \gamma^*(P_p, P_c)$ would ensure that the BER performance of the primary decoder remains unaffected with the introduction of the cognitive user. However, note that the primary decoder will calculate its channel log-likelihood ratios by assuming that the noise plus interference term is Gaussian – which might not be the case in a practical setup. Thus even though the SNR remains the same, the BER performance of the primary decoder will not be the same as that without the cognitive user. Fortunately, our simulations indicate that the dirty-paper coded output $\tilde{X}_c[n]$ is close to Gaussian and the performance of the primary decoder is not adversely affected by assuming that the interference is Gaussian. Because of this reason, we always choose $\gamma = \gamma^*(P_p, P_c)$ in our coding setup.

---

[1]In general, the log-likelihood ratios for LDPC decoding with high-level modulation formats cannot be written as functions of the received SNR alone.

1. Performance limits

With the choice of $\gamma = \gamma^*(P_p, P_c)$, the information theoretic achievable rate for the primary user is given as

$$R_p = \frac{1}{2}\log\left(1 + P_p\right). \tag{6.3}$$

On the other hand, information theoretically, the cognitive user does not see any interference from the primary user transmission because of DPC. Thus, the achievable rate for the cognitive user is given as

$$R_c = \frac{1}{2}\log\left(1 + (1-\gamma^*)P_c\right). \tag{6.4}$$

It was shown in [104], that the rates in (6.3) and (6.4) define the capacity region in the low power regime. In other words, for the case when $a < 1$, these rates are the best that can be achieved. For a given rate requirement R_p at the primary user, the theoretical minimum for the primary user power is simply given as

$$P_{p,\min} = 2^{2R_p} - 1. \tag{6.5}$$

Using some trivial algebra, it can be shown that the received SNR at the secondary base station given by $\big(1 - \gamma^*(P_p, P_c)\big)P_c$ is an increasing function of the cognitive user power P_c. Hence, for a given rate requirement R_c for the cognitive user, the theoretical minimum for the cognitive user power is given as

$$P_{c,\min} = \{P_c | \big(1 - \gamma^*(P_p, P_c)\big)P_c = 2^{2R_c} - 1\}. \tag{6.6}$$

D. Dirty-paper Coding Scheme

Several research groups have focused on designing practical DPC schemes, e.g. [34, 33, 35]. Since DPC is a source and channel coding problem, the DPC encoder contains

a channel as well as source coding component. Focusing on the high rate regime, [33] proposed a dirty-paper coding scheme based on nested turbo codes. Because of the nested nature of the scheme, the presence of the random interleaver in the turbo-channel code negatively impacts the performance of the TCQ source code. Whereas this does not affect the overall performance too much at higher transmission rates, the impact on lower rates is much more pronounced. Indeed, using a 256-state TCQ as the source code, the scheme in [33] performs only 1.42 dB from capacity at a transmission rate of 1.0 b/s, whereas the gap to capacity increases to 2.65 dB at a lower transmission rate of 0.5 b/s. On the other hand, the schemes in [34, 35] employ IRA codes as the channel code and TCQ as the source code. As opposed [33], these schemes do not suffer loss in source coding performance and were shown to perform near capacity at low transmission rates of 0.25 b/s. However, in their original format, these schemes cannot achieve a rate higher than 1.0 b/s. In this work, we use the coding framework of the scheme in [34] and adapt it to higher transmission rates. Our extension to higher rates is analogous to the extension of an LDPC code with BPSK modulation to that over a higher order constellation using multilevel coding with multi-state decoding [108]. The motivating factors behind using an LDPC code based DPC scheme over a turbo-coded scheme are:

1). Since the source code does not suffer any degradation in performance (as the scheme in [33] does), we are able to reduce the gap to capacity. At a transmission rate of 1.0 b/s, a block length $L = 50,000$ and a 256-state TCQ as the source code, we are able to achieve a gap of 0.78 dB to capacity, as opposed to 1.42 dB in [33] with the same block length and a 256-state TCQ.

2). Since the basic coding framework for our high-rate scheme remains the same as that of the low-rate scheme of [34], an obvious advantage is that switching between the low rate and the high rate regime does not require the coding setup to change.

This is particularly attractive in situations where the transmission rate needs to be adapted to the instantaneous channel quality.

In the following, we give the details of our proposed DPC scheme by discussing the encoding and decoding separately.

1. Encoding

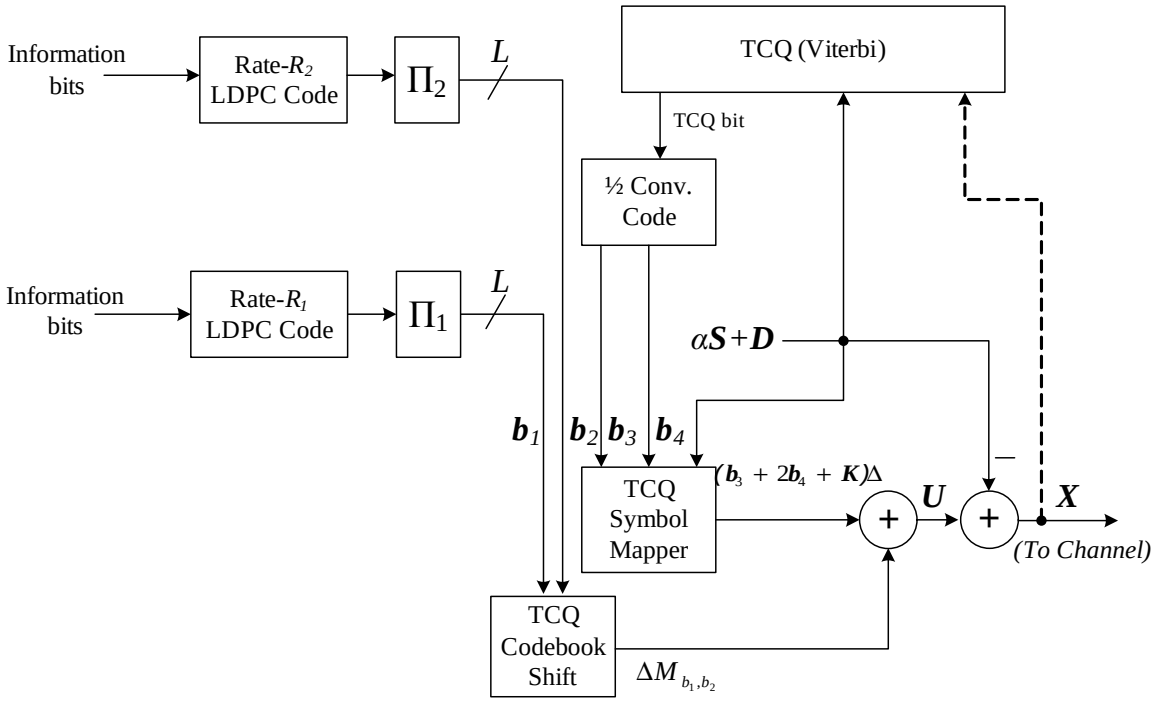


Fig. 47. Proposed DPC encoder with two levels

The encoding scheme for our DPC scheme with two levels is shown in Fig. 47, where we use LDPC codes and TCQ as the channel and source coding components, respectively. In the following, we briefly discuss the two coding components. Although all our discussions below are specific to a DPC scheme with two levels, we point out that the same design methodology can be easily extended to the case with more than

two levels.

a. Channel coding

For the channel coding component, the length k message $\mathbf{m}_c$ is split into two parts. The first part of length LR_1 is encoded using a rate-R_1 LDPC code. The LDPC code is characterized by the variable node degree distribution (from the edge perspective) $\lambda_1(x) = \sum_{d=2}^{D_{1v}} \lambda_{1d} x^{d-1}$, where λ_{1d} is the fraction of edges connected to a degree d variable node, and D_{1v} is the maximum variable node degree. On the other hand, the check node degree distribution (from the edge perspective) is given as $\rho_1(x) = \sum_{d=1}^{D_{1c}} \rho_{1d} x^{2-1}$, where D_{1c} is the maximum check node degree. Similarly, the second part of the information sequence of length LR_2 is encoded using a rate-R_2 LDPC code characterized by degree distributions $\lambda_2(x)$ and $\rho_2(x)$. The coded bit sequences from the two LDPC codes, each of length L, are first randomly interleaved which is required, at least in principle, to facilitate iterative decoding between the source and channel code at the destination. Let the output of these interleavers be denoted by $\mathbf{b}_1$ and $\mathbf{b}_2$, as shown in Fig. 47. At any given time instance, elements from the bit streams $\mathbf{b}_1$ and $\mathbf{b}_2$ are used to select one out of four TCQ codebooks, as will be explained in the source coding part. The overall rate of the dirty-paper code can be calculated as $R = k/L = R_1 + R_2$ b/s. Note that with the two levels discussed here, the maximum achievable rate is 2 b/s as opposed to 1 b/s for the schemes of [34, 35]. As mentioned earlier, the same methodology can be used to devise schemes for even higher transmission rates by increasing the number of levels. In general, a maximum rate of P b/s can be achieved by using P levels.

b. Source coding

As mentioned in Chapter II, a DPC scheme should quantize the interference and transmit the error sequence, with the quantization codebook selected by the channel codeword. With an ideal quantization source code, this approach is known to achieve capacity on the dirty-paper channel [35]. Thus, even though DPC is inherently a channel coding problem, the presence of the interference entails the need for a strong source coding/quantization element.

For practical source coding, we employ TCQ, the strongest quantization code known in the literature. The input to be quantized is $\mathbf{V} = \alpha\mathbf{S} + \mathbf{D}$, where α is Costa's MSE scaling factor [31] given by $\alpha = \frac{SNR}{1+SNR}$, $\mathbf{S}$ is the known interference/side-information sequence, and $\mathbf{D}$ is the random dither shared by the encoder and the decoder, and is required to make the quantization error independent of the quantized output – a condition necessary for achieving capacity on the dirty-paper channel [35]. The basic component of the TCQ code is a rate-1/2 convolutional code, which outputs the bit streams $\mathbf{b}_3$ and $\mathbf{b}_4$. At a time instance n, $n = 1, \ldots, L$, we let the channel coded bits $b_1[n]$ and $b_2[n]$ select one out of four TCQ codebooks to which the convolutional code outputs are mapped. Since the interference can have any arbitrary variance, each one of the four TCQ codebooks is replicated infinitely in both directions. Amongst the replicated copies of the codebook, the one closest to $V[n]$ is chosen. Since the performance of the underlying TCQ quantization codebook is independent of any shift of the codeword, we let the four TCQ codebooks be shifted versions of each other, as indicated by Fig. 47. Mathematically, the shifted TCQ symbol at time n as a function of the bits $b_i[n]$, $i = 1, 2, 3, 4$ and $V[n]$ is given by (we omit the time indices of the bits $b_i[n]$, $i = 1, 2, 3, 4$ for notational convenience)

$$U[n] = f(b_1, b_2, b_3, b_4, V[n]) = \big[(b_3 + 2b_4) + M_{b_1,b_2} + 4K[n]\big]\Delta, \tag{6.7}$$

where Δ is the step size of the TCQ codebook, $M_{b_1,b_2} \in \mathbb{R}$ is the relative shift of the TCQ codebook as a function of the channel coded bits b_1 and b_2, and $K[n]$ is indicative of the fact that the codebooks are replicated infinitely and the copy closest to $V[n]$ is selected. Mathematically,

$$K[n] = \arg\min_{k\in\mathbb{Z}} \big|(b_3 + 2b_4)\Delta + M_{b_1,b_2}\Delta + 4k\Delta - V[n]\big|^2. \tag{6.8}$$

The mapping of the bits $b_i[n]$, $i = 1, 2, 3, 4$ to the output symbol is graphically illustrated in Fig. 48. The symbols represented by the circles correspond to the basic TCQ codebook (corresponding to $k = 0$) when the channel coded bits are $[b_1, b_2] = [0, 0]$. Similarly, the squares correspond to $[b_1, b_2] = [1, 0]$, diamonds to $[b_1, b_2] = [0, 1]$ and triangles to $[b_1, b_2] = [1, 1]$. As can be seen, the TCQ codebooks corresponding to different channel coded bits are only shifted versions of each other. Within each codebook, the constellation points are uniformly spaced with the step size being Δ.

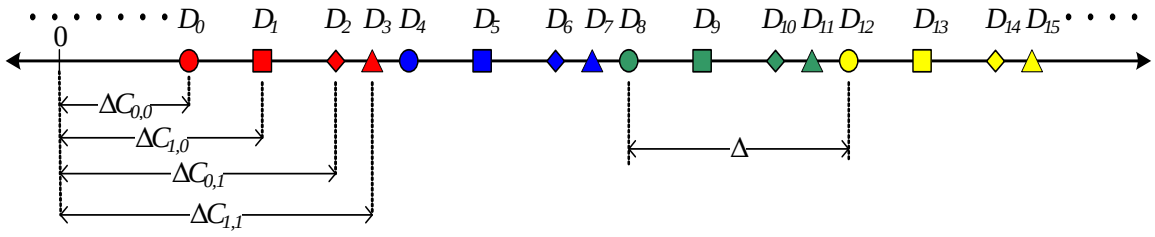


Fig. 48. The basic TCQ codebook which is repeated infinitely in both directions. The bits b_1, b_2, b_3 and b_4 are mapped to the output symbol $D_{\sum_{i=1}^{4} b_i 2^{i-1}}$.

Given the channel coded bit streams $\mathbf{b}_1$ and $\mathbf{b}_2$, the quantization procedure

involves choosing the bit streams $\mathbf{b}_3$ and $\mathbf{b}_4$ such that the overall MSE is minimized, i.e.

$$\left[\mathbf{b}_3, \mathbf{b}_4\right] = \arg \min_{\left[\mathbf{b}_3,\mathbf{b}_4\right] \in \mathcal{C}} \frac{1}{L} \sum_{n=1}^{L} \left[f(b_1[n], b_2[n], b_3[n], b_4[n], V[n]) - V[n]\right]^2, \quad (6.9)$$

where $\left[\mathbf{b}_3, \mathbf{b}_4\right] \in \mathcal{C}$ is indicative of the constraint that the bit streams should form valid codeword of the convolutional code. The optimization problem in (6.9) can be solved using the Viterbi algorithm. Finally, the error sequence $\mathbf{X} = \mathbf{U} - \mathbf{V}$ is transmitted over the channel. Note that the transmitter power is in fact the overall mean-square error (MSE). Thus for a given constraint on the transmit power, one needs to search for the TCQ step size Δ for which the MSE is equal to the required transmission power.

2. Decoding

The DPC decoding scheme is shown in Fig. 49. The received sequence is first scaled by α followed by dither removal. Thus the output after the dither removal is

$$\mathbf{Y}' = \alpha \mathbf{Y} + \mathbf{D},$$

which can be equivalently written as [35]

$$\mathbf{Y}' = \mathbf{U} + \mathbf{Z}',$$

where $\mathbf{Z}'$ is the equivalent Gaussian noise independent of $\mathbf{U}$ [35]. If P_Z is the variance of AWGN on the actual transmission channel, the variance of Z' is given by αP_Z [35]. The decoding is done by iterative message passing between the BCJR algorithm (on the TCQ convolutional code trellis) and the belief propagation (BP) algorithms on the decoding graph of the two LDPC codes. The BCJR algorithm evaluates the

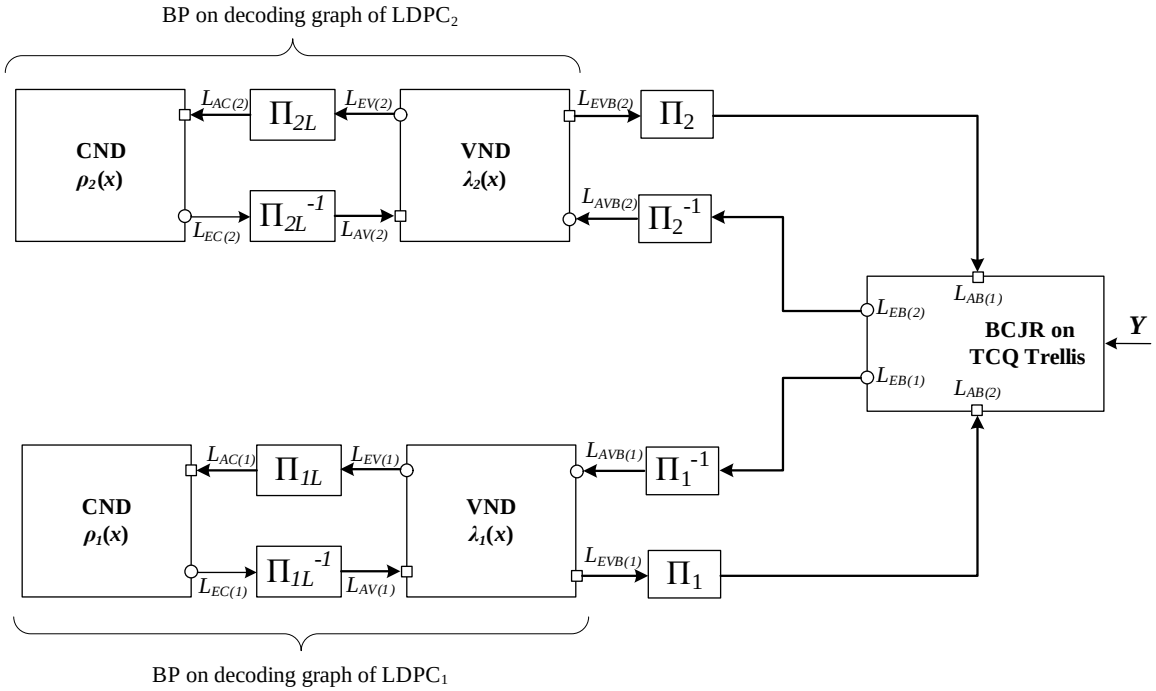


Fig. 49. DPC decoder.

extrinsic log-likelihood ratios (LLRs) on the coded data sequences $\mathbf{b}_1$ and $\mathbf{b}_2$ (input to the TCQ) with the respective a-priori LLRs coming from the LDPC decoding graphs after interleaving (these a-priori LLRs are initialized to zero for the first iteration) and with $\mathbf{Y}'$ as the channel input. The extrinsic LLRs from the BCJR algorithm are first de-interleaved and then fed into the BP algorithms as the a-priori LLRs. We consider two schedules for decoding, serial and parallel. We briefly describe the two decoding schedules in the following.

Serial Decoding Schedule: In the serial decoding schedule (SDS), we first perform iterative message massing between the BCJR on the TCQ trellis, and the BP decoder for the first LDPC decoder. During this message passing, the a-priori LLRs $\mathbf{L}_{AB(2)}$ to the BCJR are treated as zeros. This iterative message passing continues, until

some stopping criterion is met[2]. We then use the extrinsic LLRs from the first LDPC decoding graph, and run iterative message passing between the BCJR algorithm and the decoding graph of the second LDPC code. The SDS is summarized in the following algorithm (see Table VIII for a notation of different LLRs)

1: Initialize: $\mathbf{L}_{AV(i)} = \mathbf{0}$, $\mathbf{L}_{AB(i)} = \mathbf{0}$, for $i = 1, 2$

2: **While** (stopping criterion not met) % *Message passing for LDPC*$_1$

3: Run BCJR on TCQ trellis with $\mathbf{L}_{AB(i)}$ as input. Outputs $\mathbf{L}_{EB(i)}$, $i = 1, 2$.

4: Assign: $\mathbf{L}_{AVB(1)} = \Pi_1^{-1}\left(\mathbf{L}_{EB(1)}\right)$.

5: Run $\mathcal{I}$ iterations of BP algorithm on decoding graph of LDPC$_1$ with $\mathbf{L}_{AVB(1)}$ as input channel LLR, and $\mathbf{L}_{AV(1)}$ as the CND to VND LLRs. Output extrinsic LLRs $\mathbf{L}_{EVB(1)}$ ($\mathbf{L}_{AV(1)}$ is also changed).

6: Assign: $\mathbf{L}_{AB(1)} = \Pi_1\left(\mathbf{L}_{EVB(1)}\right)$.

7: **end while**

8: Repeat the message passing described by the while loop above, but with decoding graph of LDPC$_2$ instead of LDPC$_1$. Keep $\mathbf{L}_{AB(1)}$ fixed to values obtained from message passing with decoding graph of LDPC$_1$.

9: Assign: $\mathbf{L}_i = \mathbf{L}_{EB(i)} + \mathbf{L}_{EVB(i)}$, $i = 1, 2$. Hard-threshold $\mathbf{L}_i$ to obtain decisions.

Parallel Decoding Schedule: In the parallel decoding schedule PDS, we run iterative BP on the two decoding graphs simultaneously. The PDS is summarized below.

[2]In our simulations, we let the iterations continue until a maximum number of iterations is reached, or when a valid codeword has been decoded.

Table VIII. Summary of LLR notations used in the DPC decoding.

Notation	**Description of LLRs**
$\mathbf{L}_{AV(i)}$	LLRs from CND to VND at LDPC$_i$
$\mathbf{L}_{EVB(i)}$	Extrinsic LLRs from VND of LDPC$_i$ to BCJR
$\mathbf{L}_{AVB(i)}$	Channel LLRs inputs to VND of LDPC$_i$ (from BCJR decoder)
$\mathbf{L}_{AB(i)}$	A-priori LLRs to the BCJR decoder for bit stream $\mathbf{b}_i$
$\mathbf{L}_{EB(i)}$	Extrinsic LLRs from the BCJR decoder for bit stream $\mathbf{b}_i$

1: Initialize: $\mathbf{L}_{AV(i)} = \mathbf{0}$, $\mathbf{L}_{AB(i)} = \mathbf{0}$, for $i = 1, 2$

2: **While** (stopping criterion not met)

3: Run BCJR on TCQ trellis with $\mathbf{L}_{AB(i)}$ as input. Outputs $\mathbf{L}_{EB(i)}$, $i = 1, 2$.

4: Assign: $\mathbf{L}_{AVB(i)} = \Pi_i^{-1}\left(\mathbf{L}_{EB(i)}\right)$, $i = 1, 2$.

5: Run $\mathcal{I}$ iterations of BP algorithm on decoding graph of LDPC$_i$ with $\mathbf{L}_{AVB(i)}$ as input channel LLR, and $\mathbf{L}_{AV(i)}$ as the CND to VND LLRs. Output extrinsic LLRs $\mathbf{L}_{EVB(i)}$ ($\mathbf{L}_{AV(i)}$ is also changed). $i = 1, 2$.

6: Assign: $\mathbf{L}_{AB(i)} = \Pi_i\left(\mathbf{L}_{EVB(i)}\right)$, $i = 1, 2$.

7: **end while**

8: Assign: $\mathbf{L}_i = \mathbf{L}_{EB(i)} + \mathbf{L}_{EVB(i)}$, $i = 1, 2$. Hard-threshold $\mathbf{L}_i$ to obtain decisions.

The merits/demerits of the two decoding schedules will be discussed in subsequent sections.

3. DPC Design

As mentioned earlier, the DPC scheme consists of a source coding and a channel coding component. The design for the source code involves choosing appropriate shifts M_{b_1,b_2}, $b_1, b_2 \in \{0,1\}$ of the TCQ codebook, whereas the channel code design involves choosing appropriate degree distributions for the LDPC codes. In the following, we discuss these two aspects separately.

a. Source code design

The relative shifts M_{b_1,b_2}, $b_1, b_2 \in \{0,1\}$ of our scheme are analogous to the TCQ codebook shift mentioned in [34]. Since the scheme of [34] has only one channel code, there are only two shift parameters as opposed to four in our case. We follow the same approach as in [34] for selecting the appropriate shifts. The shifts characterize the TCQ induced channel and control the amount of information the induced channel conveys about the original message. This information can be written as

$$\begin{aligned} C\left(\{M_{b_1,b_2}\}\right) &\equiv \frac{1}{L} I\left(\mathbf{b}_1, \mathbf{b}_2; \mathbf{Y}' \middle| \{M_{b_1,b_2}\}\right) \\ &= \frac{1}{L} I\left(\mathbf{b}_1; \mathbf{Y}' \middle| \{M_{b_1,b_2}\}\right) + \frac{1}{L} I\left(\mathbf{b}_2; \mathbf{Y}' | \mathbf{b}_1, \{M_{b_1,b_2}\}\right) \qquad (6.10) \\ &\equiv C_1\left(\{M_{b_1,b_2}\}\right) + C_2\left(\{M_{b_1,b_2}\}\right), \qquad (6.11) \end{aligned}$$

where (6.10) follows from the chain rule of mutual information. In order to evaluate the information $C\left(\{M_{b_1,b_2}\}\right)$ as a function of the relative shifts $\{M_{b_1,b_2}\}$, we utilize the EXIT chart strategy and the area theorem [64], according to which, if $I_E(I_A)$ is the extrinsic information conveyed by any decoder as a function of the a-priori information I_A, the capacity on that channel can be approximated by the area under

the curve and is specifically given as

$$\mathcal{A} = \int_0^1 I_E(I) dI. \tag{6.12}$$

Although the expression (6.12) is exact only for a binary erasure channel, it is known to approximate the capacity well on other channels, even if the channel has memory (such as the TCQ induced channel). Thus, this methodology should be able to provide a good approximation to the channel capacity in (6.11).

In order to generate the EXIT curve, we resort to Monte-Carlo simulations. We first generate the bit streams $\mathbf{b}_1$ and $\mathbf{d}_2$ randomly, each of length $L = 100,000$. We also generate the length-L side-information sequence $\mathbf{S}$ randomly. We assume that the transmissions from the primary user are mapped to a 4-PAM constellation. Since the primary user transmissions act as the side-information sequence to the DPC encoder at the cognitive user, we make sure that the side-information is randomly drawn from a 4-PAM constellation. For some given shifts M_{b_1,b_2}, $b_1, b_2 \in \{0,1\}$, the sequence $\mathbf{V} = \alpha\mathbf{S} + \mathbf{D}$ (the dither $\mathbf{D}$ is uniformly distributed over the interval $[0, 4\Delta]$) is quantized using the TCQ code, as mentioned in Section D-1. The step-size Δ is chosen such that the resulting MSE is equal to the required received SNR of $(1-\gamma)P_c$ at the cognitive user receiver. The error sequnce is then transmitted over an AWGN channel with unit noise variance. Next, we run the BCJR algorithm on the TCQ trellis, with $\mathbf{Y}' = \alpha\mathbf{Y} + \mathbf{D} \approx \mathbf{U} + \mathbf{Z}'$ as the channel input, and $\mathbf{L}_{AB(1)}$ and $\mathbf{L}_{AB(2)}$ as the a-priori LLRs for the sequence $\mathbf{b}_1$ and $\mathbf{b}_2$, respectively. In order to evaluate $C(\{M_{b_1,b_2}\})$, we follow the chain rule of mutual information and evaluate $C_1(\{M_{b_1,b_2}\})$ and $C_2(\{M_{b_1,b_2}\})$ in (6.11) separately. First, for evaluating $C_1(\{M_{b_1,b_2}\})$, we assume, as (6.10) dictates, that there is no a-priori information available regarding the bit-sequence $\mathbf{b}_2$, i.e. we assume that $\mathbf{L}_{AB(2)} = \mathbf{0}$. For a given a-priori information $I_{AB(1)}$ to the BCJR decoder on the bit sequence $\mathbf{b}_1$, the a-priori LLR sequence $\mathbf{L}_{AB(1)}$ evaluated

at position n is generated by

$$L_{AB(1)}[n] = (1 - 2b_1[n])J^{-1}(I_{AB}) + Z[n], \qquad n = 1, \ldots, L, \tag{6.13}$$

where Z[n] is a zero-mean Gaussian random variable with variance $2J^{-1}(I_{AB})$, and where $J(\mu)$, as defined in the previous chapters, is the information a Gaussian LLR of mean μ and variance 2μ conveys about the message bit[3]. The BCJR algorithm outputs the extrinsic LLR sequences $\mathbf{L}_{EB(1)}$. The extrinsic information on the bit stream $\mathbf{b}_1$ can be evaluated as

$$I_{EB(1)} = \frac{1}{L}\sum_{n=1}^{L} \log_2 \left(\frac{2}{1 + \exp\left((2b_1[n] - 1)L_{e(1)}[n]\right)} \right). \tag{6.14}$$

By repeating the procedure mentioned above for several values of $I_{AB(1)}$, one can obtain the $I_{EB(1)}$ versus $I_{AB(1)}$ curve which can then be used to approximate the capacity $C_1\left(\{M_{b_1,b_2}\}\right)$ of the TCQ induced channel by (6.12).

In order to evaluate $C_2\left(\{M_{b_1,b_2}\}\right)$, we have to assume, as indicated in (6.10), complete a-priori information about the bit sequence $\mathbf{b}_1$. Thus, using $L_{AB(1)}[n] = (1 - 2b_1[n]) \times \infty$, we run the BCJR algorithm on the TCQ trellis for a given a-priori information $I_{AB(2)}$ and the corresponding LLR sequence $\mathbf{L}_{AB(1)}$. The extrinsic information $I_{EB(2)}$ is evaluated from the extrinsic LLRs in the same manner as in (6.14) and the approximated capacity $C_2\left(\{M_{b_1,b_2}\}\right)$ is evaluated by calculating the area under the $I_{EB(2)}$ versus $I_{AB(2)}$ curve.

Finally, we point that in order maximize the capacity of the TCQ induced capacity, one needs to perform an exhaustive search over $\{M_{b_1,b_2}\}$. In order to illustrate

[3]The TCQ induced channel might not be symmetric, and hence, in general the usual assumption of the symmetry of the LLRs might not be valid. However, we use the concept of i.i.d. channel adapters [109] using which the channel can be transformed into an equivalent symmetric channel, and hence the assumption of the LLRs having a symmetric distribution becomes valid.

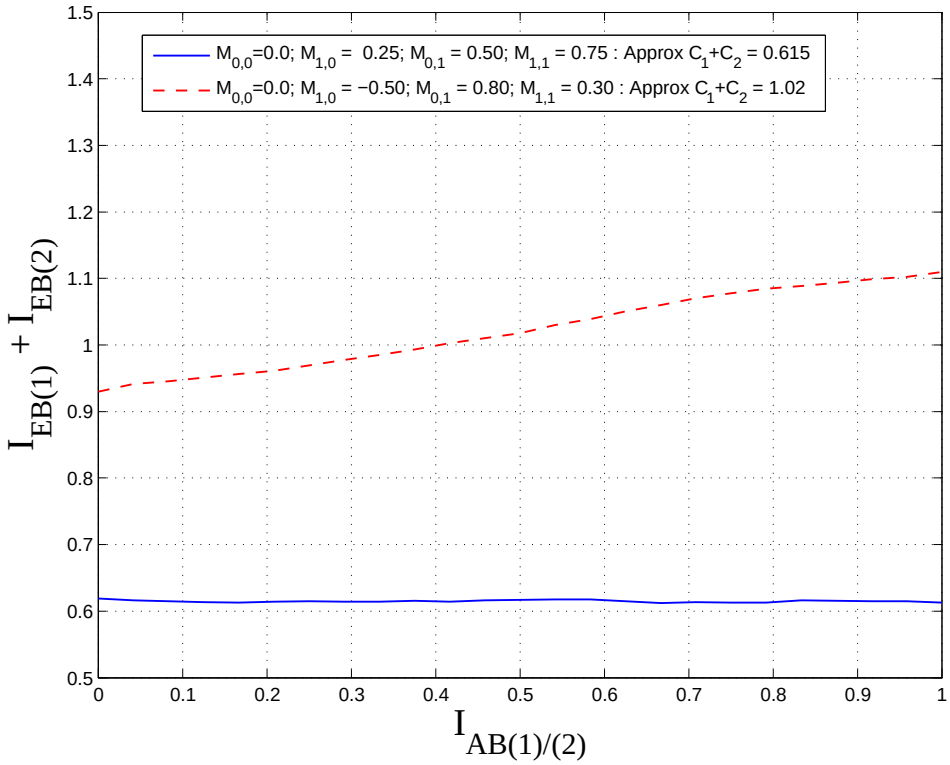


Fig. 50. EXIT curves for the BCJR decoder for two sets of shifts of the TCQ codebook. The SNR is fixed at 5.2 dB and the TCQ uses a 256-state feedback convolutional code defined by the polynomials $h_0 = 625$ and $h_1 = 242$ in octal form. The approximate capacities are evaluated using (6.12).

the huge difference that an appropriate choice of the shifts can make, we shown the EXIT curves for two sets of shifts in Fig. 50. The SNR is fixed at 5.2 dB and the TCQ uses a 256-state feedback convolutional code defined by the polynomials $h_0 = 625$ and $h_1 = 242$ in octal form. It can be observed that when employing uniform shifts, i.e., when $\{M_{0,0}, M_{1,0}, M_{0,1}, M_{1,1}\} = \{0.00, 0.25, 0.50, 0.75\}$, the area under the curve can be approximated as $C = 0.615$ (the individual areas are given as C_1=0.098 and C_2=0.517). On the other hand, when $\{M_{0,0}, M_{1,0}, M_{0,1}, M_{1,1}\} = \{0.00, -0.50, 0.80, 0.30\}$, the approximated capacity is $C = 1.02$ (the individual areas

under the curve are $C_1 = 0.196$ and $C_2 = 0.824$). Thus the second set of shifts gives a gain of more than 0.4 bits/sample over the uniform shifts of the first set. In Fig. 51, we show the individual EXIT curves corresponding to the two bit streams $\mathbf{b}_1$ and $\mathbf{b}_2$.

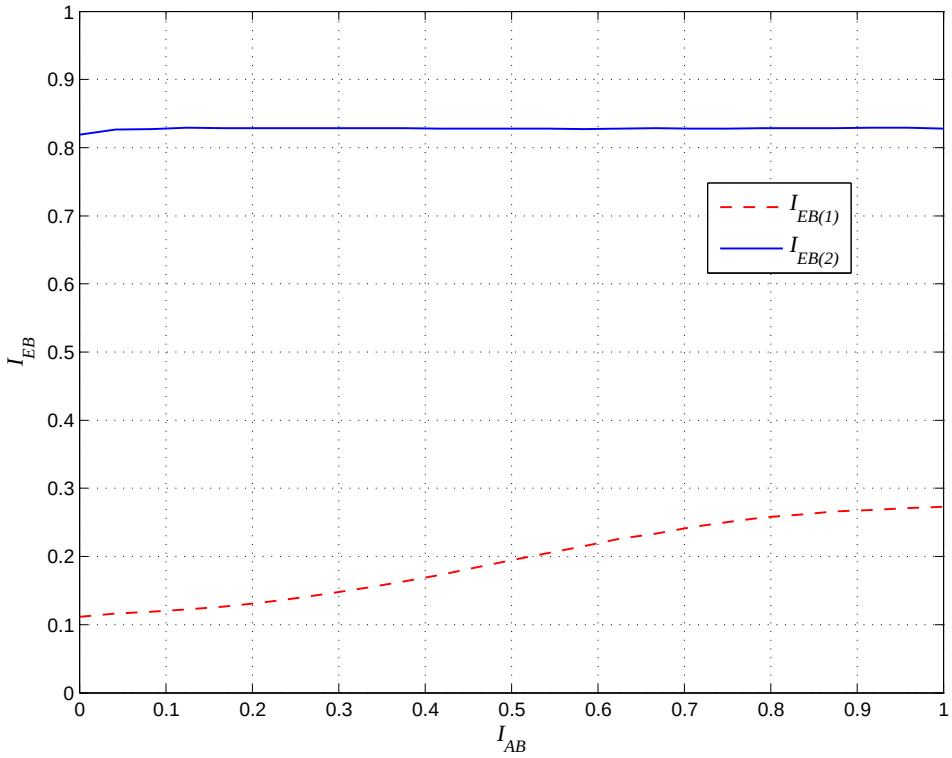


Fig. 51. EXIT curves of the BCJR decoder corresponding to the two bit streams $\mathbf{b}_1$ and $\mathbf{b}_2$. The SNR is fixed at 5.2 dB and the TCQ uses a 256-state feedback convolutional code defined by the polynomials $h_0 = 625$ and $h_1 = 242$ in octal form. The TCQ codebook shifts are set at $\{M_{0,0}, M_{1,0}, M_{0,1}, M_{1,1}\} = \{0.00, -0.50, 0.80, 0.30\}$.

b. Channel code design

As in the previous subsection, we once again use the Gaussian assumption and the EXIT chart strategy to design the degree distributions for the LDPC codes. We fix the variable node degree profiles $\lambda_1(x)$ and $\lambda_2(x)$ and optimize the check node profiles $\rho_1(x)$ and $\rho_2(x)$. We point out that in most works in the literature, the variable node degree profiles are optimized for fixed check node profiles. However, as we will see in the following, having the reverse for our DPC design results in a simpler optimization problem which can be solved using linear programming. At the same time, this design methodology promises good performance, as will be seen in Section E.

For the design of the degree distributions, we assume an asymptotically large block length L. The chain rule of mutual information in (6.10) gives us an insight into how the design procedure can be simplified: One can first design the degree distribution $\rho_1(x)$ for the first LDPC code while assuming that the a-priori information to the BCJR decoder from the second LDPC code is zero. Once $\rho_1(x)$ has been designed to guarantee error free recovery of $\mathbf{b}_1$, one can then design $\rho_2(x)$ with the a-priori information to the BCJR decoder from the first LDPC code being one. A keen reader would observe that this strategy corresponds to the SDS presented in Section D-2. Thus, the SDS is in fact information theoretically optimum – the design of the degree distributions is carried out with the SDS in mind. However, we will use the degree distributions designed in this section to implement a SDS as well as PDS in Section E. As will be seen, the degree distributions perform very well with a PDS as well, even though they are designed for a SDS. In fact, PDS outperforms SDS by about 0.1 dB.

As mentioned before, we first design the check node degree distribution $\rho_1(x)$ assuming the variable node degree distribution $\lambda_1(x)$ is fixed. The information flow

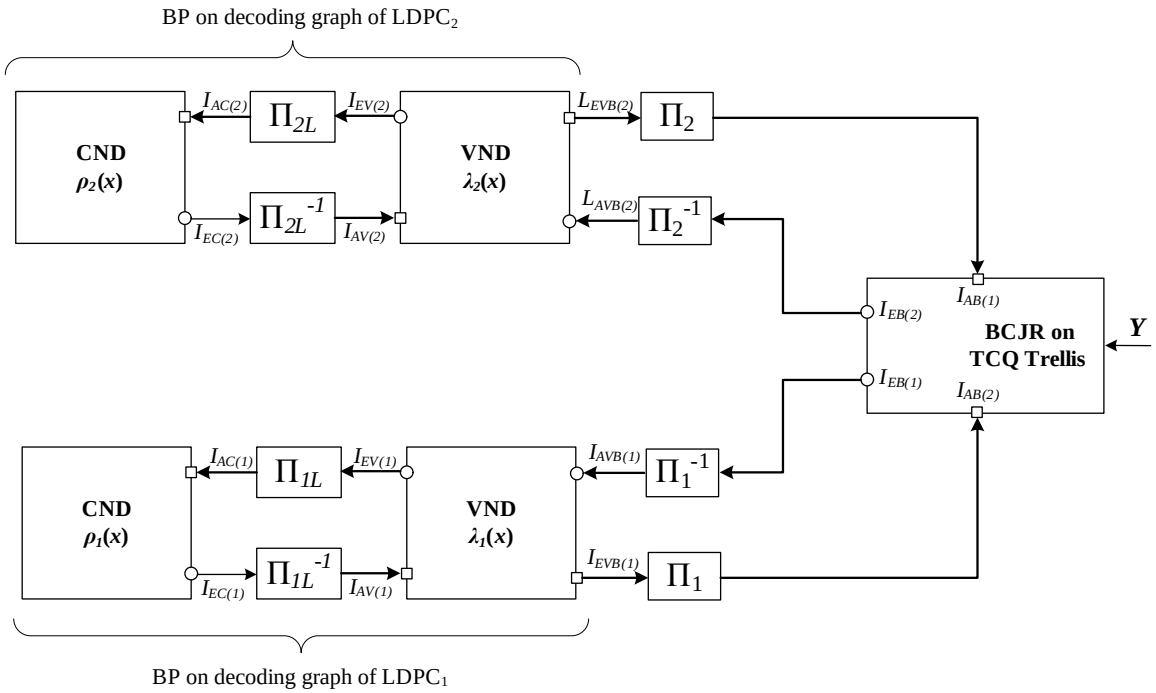


Fig. 52. Information flow for the channel code design.

for the DPC decoder is indicated in Fig. 52. The figure is the same as Fig. 49 except that the LLRs have been replaced with the respective informations – we reproduce the figure here for the reader's convenience. Let the a-priori information to the variable nodes of the first LDPC code be given as $I_{AV(1)}$ with $I_{AV(1)} \in [0, 1]$. Assuming that the a-priori LLR to the variable nodes corresponding to this information is Gaussian[4] with mean $J^{-1}\left(I_{AV(1)}\right)$, and variance $2J^{-1}\left(I_{AV(1)}\right)$, we evaluate the probability mass function (pmf) $\mathcal{P}_{V(1)\to B}$ of the extrinsic LLRs from the variable nodes of LDPC_1 to the BCJR decoder using discretized density evolution [66]. We now need to evaluate the information transfer function at the BCJR decoder, for which we once again resort to Monte-Carlo simulations. We first generate two blocks of equally probable

[4]Once again, using the concept of i.i.d. channel adapters [109], we can assume the transmission of the all-zero codeword and symmetry of LLR distributions.

i.i.d. channel adapters $\mathbf{c}_1$ and $\mathbf{c}_2$. Then encoding and transmission over the AWGN channel is simulated as in the source code design with $\mathbf{b}_1 \oplus \mathbf{c}_1$ and $\mathbf{b}_2 \oplus \mathbf{c}_2$ treated as the input sequence to the TCQ codebook mapper, where $\oplus$ represents bit-wire xor. As pointed out in [109], the modified channel with these channel adapters becomes symmetric and hence one can assume that $\mathbf{b}_1$ and $\mathbf{b}_2$ are the all-zero sequences. Thus the inputs to the TCQ codebook mapper are simply $\mathbf{c}_1$ and $\mathbf{c}_2$. We then generate a block of intermediate a-priori LLRs $\mathbf{L}'_{AB(1)}$ according to the pmf $\mathcal{P}_{V(1)\to B}$ evaluated earlier using density evolution. The a-priori LLRs to the BCJR algorithm are given as

$$L_{AB(1)}[n] = \big(1 - 2c_1[n]\big) L'_{AB(1)}[n] \quad n = 1, \ldots, L.$$

Applying the BCJR decoder with $\mathbf{L}_{AB(1)}$ and $\mathbf{L}_{AB(2)} = \mathbf{0}$ as a-priori LLRs, and the equivalent channel output $\mathbf{Y}'$, we then obtain the intermediate LLRs $\mathbf{L}'_{EB(1)}$ for the combined values of the codeword bits and the channel adapters. The LLRs for the codeword bits are evaluated by reversing the effect of the channel adapters. Specifically,

$$L_{EB(1)}[n] = \big(1 - 2c_1[n]\big) L'_{EB(1)}[n] \quad n = 1, \ldots, L.$$

Using the sequence $\mathbf{L}_{EB(1)}$, we find the histogram to evaluate the pmf $\mathcal{P}_{B\to V(1)}$, i.e., we evaluate the pmf of the LLRs going from the BCJR decoder to the variable nodes of LDPC$_1$. Then with $\mathcal{P}_{B\to V(1)}$ as the pmf of the channel LLR, and $\mathcal{P}_{AV} = \mathcal{N}\left(J^{-1}\left(I_{AV(1)}\right), 2J^{-1}\left(I_{AV(1)}\right)\right)$ as the pmf of the a-priori LLR from the check nodes, we evaluate, as a function of the information I_{AV}, the pmf $\mathcal{P}_{EV}\left(I_{AV}\right)$ of the extrinsic LLR from the variable node to the check node using density evolution. Note that this pmf can be easily evaluated since the variable node degree distribution $\lambda_1(x)$ is assumed to be fixed, and is hence not a design variable.

Employing density evolution, we use the extrinsic pmf $\mathcal{P}_{EV}$ as the a-priori infor-

mation and evaluate the extrinsic information at check node of degree d as a function of I_{AV} as $\mathcal{P}_{EC}^{d}\left(I_{AV}\right)$. This pmf is then mapped to an extrinsic information term $I_{EC}^{d}\left(I_{AV}\right)$ as

$$I_{EC}^{d}\left(I_{AV}\right)=\sum_{l}\mathcal{P}_{EC}^{d}\left(I_{AV}\right)[l]\log_2\left(\frac{2}{1+\exp(-l)}\right), \tag{6.15}$$

where $\mathcal{P}_{EC}^{d}\left(I_{AV}\right)[l]$ is the probability that the LLR is equal to the discretized value l. The overall extrinsic information from the check node is then given as

$$I_{EC}\left(I_{AV}\right)=\sum_{d=1}^{D_{1c}}\rho_d I_{EC}^{d}\left(I_{AV}\right). \tag{6.16}$$

For convergence of BP decoding (in the Gaussian assumption and EXIT function sense), the following constraint should be satisfied for all $I_{AV}\in[0,1)$ [64]

$$\sum_{d=1}^{D1c}\rho_d I_{EC}^{d}\left(I_{AV}\right)>I_{AV}. \tag{6.17}$$

In addition to the convergence constraint, we require $\sum_{d=1}^{D_{1c}}\rho_d=1$, and $\rho_d>0$, $d=1,\ldots,D_{1c}$. For these constraints one should maximize the rate of the LDPC code which is equivalent to minimizing $\sum_{d=1}^{D_{1c}}\frac{\rho_d}{d}$. By discretizing the interval $I_{AV}\in[0,1)$ and requiring the constraint (6.17) to be satisfied for all values of the discretized values, the optimization problem can be easily solved using linear programming.

After the degree distribution $\rho_1(x)$ has been designed, we next design the check node degree distribution $\rho_2(x)$. For that purpose, one can follow, in principle, the same design procedure mentioned above. Since the design procedure for $\rho_1(x)$ ensures, at least in the EXIT function and Gaussian assumption sense, that the bit stream $\mathbf{b}_1$ has been perfectly decoded, one can always assume that $L_{AB(1)}[n]=(1-2c_1[n])\times\infty$ when designing $\rho_2(x)$. However, we point out a fact here that greatly simplifies the optimization of the degree distributions of LDPC_2. One can observe from Fig. 51 that the EXIT curve $I_{EB(2)}$ for the bit-stream $\mathbf{b}_2$ remains almost flat. As a result,

one need not worry about message passing between the LDPC channel decoder and the BCJR source decoder. One can simply run BCJR decoding once, and then use the resulting extrinsic information to run SPA decoding on graph of LDPC_2 until the stopping criteria has been met. In this regard, one can simply optimize the degree distributions $\lambda_2(x)$ and $\rho_2(x)$ using well established techniques for a point to point AWGN channel [103, 64, 73, 65]. *This also simplifies the decoding process in the SDS, since one can run only a few iterations of message passing between the BCJR decoder and the decoder for LDPC$_2$, or quite possibly not run any iterations between the two at all.*

As far as the rate allocation of the two LDPC codes R_1 and R_2 are concerned, one can simply follow the above design procedures to obtain the optimized LDPC codes of certain rates for a given SNR. One can then increase/decrease the overall SNR until the sum of the two rates $R_1 + R_2$ is equal to the desired rate.

E. Simulation Results

For the primary user, we fix the transmission rate at $R_p = 1$ b/s. We use the BICM scheme with a 4-PAM modulation and a rate-1/2 LDPC code with the optimized degree profiles of [110]. Simulation results for the primary user indicate that in the absence of the cognitive user, and a block length of $L = 50,000$, the transmission power required to achieve a target BER of at least 1×10^{-5} is approximately 5.95 dB. Thus, in our setup we always fix $P_p = 5.95$ dB. For the DPC scheme, we use a 256-state feedback convolutional code for TCQ with polynomials $h_0 = 625$ and $h_1 = 242$ in octal form. The relative shifts of the TCQ codebook are fixed to the second set in Fig. 50, i.e. $M_{0,0} = 0.0$, $M_{1,0} = -0.50$, $C_{0,1} = 0.80$, and $C_{1,1} = 0.30$. As mentioned earlier, the constellation step size Δ is chosen such that the quantization error in (6.9)

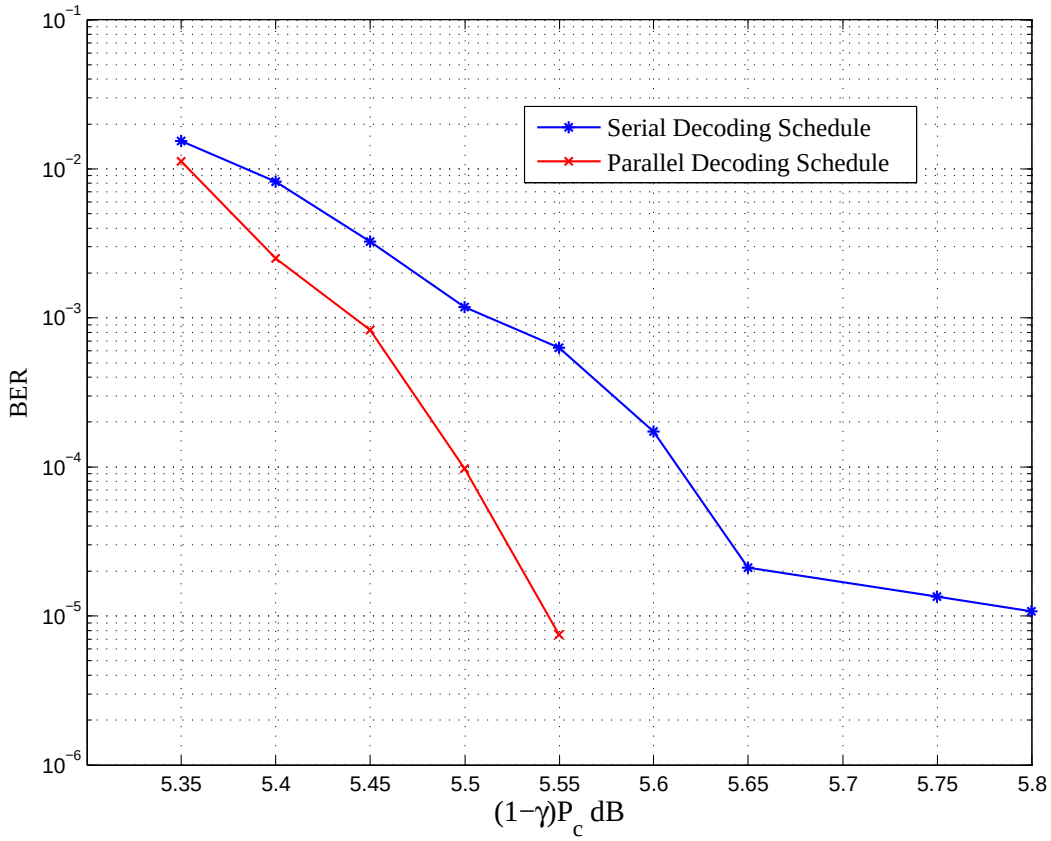


Fig. 53. Bit-error rates for the cognitive user at a transmission rate of 1 b/s, a block length of $L = 50,000$ and with a 256-state TCQ. The theoretical limit for the given rate is 4.77 dB.

is equal to the required power $(1-\gamma^*)P_c$. For the LDPC codes in the DPC scheme, the optimized degree profiles along with their rates are given in Table IX.

With the choice of the degree distributions in Table IX, the overall transmission rate for the cognitive user is $R_c = R_1 + R_2 = 1.0$ b/s. The theoretical limit for the cognitive user required for this transmission rate is evaluated from (6.6) as $P_{c,\min} = 6.67$ dB which ensures that the received SNR at the secondary receiver is $(1-\gamma^*)P_{c,\min} = 4.77$ dB. The BER for the cognitive user as a function of the received

Table IX. Optimized degree distributions for the two LDPC codes corresponding to rates $R_1 = 0.18$ and $R_2 = 0.82$. The TCQ code uses a 256-state feedback convolutional code defined by the polynomials $h_0 = 625$ and $h_1 = 242$ in octal form.

$\lambda_1(x)$	$0.19x + 0.21x^2 + 0.19x^7 + 0.2x^{31} + 0.21x^{99}$
$\rho_1(x)$	$0.0021x + 0.0721x^2 + 0.4241x^4 + 0.3575x^8 + 0.1442x^{11}$
$\lambda_2(x)$	$0.1267x + 0.1851x^2 + 0.1896x^6 + 0.0406x^7 + 0.0171x^{15} + 0.0512x^{18} + 0.0888x^{19} + 0.0299x^{20} + 0.0217x^{21} + 0.0075x^{32} + 0.0092x^{39} + 0.0079x^{53} + 0.0872x^{60} + 0.0413x^{62} + 0.0890x^{63} + 0.0071x^{68}$
$\rho_2(x)$	$0.7x^{31} + 0.3x^{32}$

SNR $(1 - \gamma^*)P_c$ at a block length of $L = 50,000$ is shown in Fig. 53 for both the SDS and PDS. As can be seen, the PDS outperforms SDS by approximately 0.1 dB, with the PDS being only 0.78 dB away from the theoretical performance limit at a BER of 1×10^{-5}. As far as we know, this is the best DPC performance reported in the literature for this rate. In addition, through our simulations we were able to verify that the performance of the single-user decoder at the primary user remains below the required 1×10^{-5} threshold even in the presence of the cognitive user, thus satisfying the requirement that the primary user remain oblivious to the presence of the cognitive user.

F. Summary

We have implemented the DPC based scheme of [104] for the CRC. We have considered the case where the cognitive radio is already aware of the primary user message and that the primary user and destination are unaware of the existence of the cog-

nitive radio. For DPC, we have proposed a scheme which employs multi-level LDPC codes for channel coding and TCQ for source coding. Using a 256-state TCQ, the scheme operates within 0.78 dB of the capacity at a transmission rate of 1.0 b/s and a block length of 50,000. Possible extensions to this work is to consider a scenario where the cognitive user does not initially know the primary user message, but knows its codebook. Hence the first phase for the cognitive user would be to listen in on the primary user transmission and attempt to decode its message before it begins its own transmissions. Such a scheme would obviously require the primary user code to be rate-compatible. Another direction of possible future research is to consider the general scheme of [106] which involves joint decoding of multiple dirty-paper codes.

CHAPTER VII

CONCLUSIONS

In this dissertation, we have considered four cooperative communication channels. For each one of these channels, we developed information theoretic coding strategies and derived the corresponding performance limits. We then implemented the coding strategies using the likes of LDPC, IRA, and Raptor codes as well as nested scalar quantization and trellis coded quantization. For each case, we find that the designed practical coding schemes operate very close to the theoretical performance limits.

The four cooperative communication channels we have studied are: (a) The Gaussian relay channel, (b) the quasi-static fading relay channel, (c) the cooperative MAC, and (d) the CRC. In the following, we provide the concluding remarks for each one of these channels.

A. The Gaussian Relay Channel

We proposed a WZ coding based CF coding strategy with BPSK modulation for the half-duplex Gaussian relay channel. As a means of implementing the WZ based CF coding, we proposed SWCNSQ and derived the corresponding achievable rates, specifically with BPSK modulation. We then proposed several simplifications to the scheme which greatly simplifies the practical implementation, but result in negligible loss in performance. Following the guiding principles from the information theoretic analysis, we then developed the first limit-approaching practical CF code design which uses LDPC codes for error correction at the source. In addition, it uses nested scalar quantization and distributed joint source-channel coding with IRA codes at the relay. We showed that once the quantization indices are recovered at the destination, the destination effectively sees the transmission from the source reaching it over two

parallel sub-channels. Using EXIT charts strategy and the Gaussian assumption, we designed good degree distributions for LDPC decoding over the two sub-channels. We simulated our design for several transmission rates, and also for different geometrical settings of the nodes. Using density evolution for asymptotically large block lengths, we found that our code design operates only 0.11 – 0.2 dB away from the information theoretic limit, whereas simulations with finite block length of 2×10^5 (and a BER of 1×10^{-5}) exhibit a gap of only 0.27 – 0.38 dB from the achievable information theoretic bound.

B. The Quasi-static Fading Relay Channel

We considered the extension of the Gaussian relay channel to the case where all links experience independent quasi-static Rayleigh fading, and where the CSI is not available at the transmitters, but is perfectly available at the destination. We considered a situation where it is essential for the destination to always decode the source message correctly, and where the network did not have any delay constraints. As a result, the system can allow the source and the relay to continue transmitting until successful decoding occurs at the destination. Under this setup, we identified rateless coding as the natural choice where each transmission from the source and/or the relay is a source of incremental redundancy. We proposed rateless coded versions of DF and CF relaying, and derived the corresponding performance limits, specifically with BPSK modulation. Since the CSI is not available at the transmitters, we proposed a novel protocol which allows an additional bit of feedback from the destination. This feedback not only serves to indicate to the source and the relay which relaying scheme to employ, but also helps informing the relay when to stop receiving and start transmitting. The proposed rateless coded protocol was then put into practice using Raptor

codes, which are not only used for traditional channel coding, but also for rateless joint source and channel coding for CF relaying. We proposed for the relaying protocol to use the same set of Raptor code degree distributions for all channel coefficients, with the degree distributions optimized so as to maximize the throughput averaged over the fading distributions of the channel coefficients. Using EXIT charts and the Gaussian assumption, we formulated the design of these degree distributions as a convex but non-linear optimization problem. Results indicate that for asymptotically large block lengths, the rateless coded relaying protocol with optimized degree distributions loses only $\sim 5\%$ in average throughput performance compared to the theoretical limit. With finite length simulations, the corresponding losses in average throughput were only $\sim 9\%$.

C. The Cooperative Multiple-access Channel

We considered cooperation in the low power regime for the MAC. The channel model assumes i.i.d. fading over all links, with the CSI not available at the transmitters. Under this channel model, we identified outage capacity as the relevant performance measure. We developed cooperation methods based on multiplexed coding in conjunction with rateless coding and found the corresponding achievable rates. In the low power regime, we used the analysis to obtain the minimum energy per bit required to achieve a certain outage probability. For cooperation, we considered two modes of operation: full duplex (CDMA), where nodes can transmit and receive simultaneously on the same frequency band, and half duplex (FDMA), where the nodes transmit and listen on different frequency bands. In order to provide a comparison of the two schemes in the low power regime, we studied their respective performances using the *outage wideband slope*. We showed that perhaps surprisingly, there is little

loss in performance when using FDMA over CDMA. This is of practical significance, since in real-time wireless systems, the former is much easier to implement than the latter. Furthermore, our results also indicated that multiplexed rateless coding comes within 0.1 dB of the upper bounds, and is hence capacity approaching. Finally, we developing practical coding methods for FDMA using multiplexed Raptor codes which operate within 0.52 and 1.1 dB of the theoretical limit for the two- and the four-user case, respectively.

D. The Cognitive Radio Channel

We considered the case of a CRC where the cognitive radio is already aware of the primary user message. In addition, we considered a situation where the primary user is oblivious to the presence of the cognitive user – the presence of the cognitive user should in no way affect the encoding/decoding process of the primary user, not should it affect its performance. Since the primary user message is known at the cognitive user, it calls for the cognitive user to dirty-paper code its message with the primary user transmissions as the side-information known at the encoder, but not at the decoder. We implemented the dirty-paper based coding strategy by proposing a DPC scheme which employs multi-level LDPC codes and TCQ as the channel and source coding component, respectively. We identified the design aspects of the source and channel code, and found that the optimized scheme operates within 0.78 dB of the channel capacity at a transmission rate of 1.0 b/s and a block length of 50,000. At this transmission rate, this is the best performance, as far as we are aware, amongst any existing DPC scheme in the literature.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I: System description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, November 2003.

[2] ——, "User cooperation diversity. Part II: Implementation aspects and performance analysis," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, November 2003.

[3] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Transactions on Information theory*, vol. 52, no. 4, pp. 1522–1544, April 2006.

[4] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 2, pp. 283–289, February 2006.

[5] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Transactions on Communications*, vol. 52, no. 9, pp. 1470–1476, September 2004.

[6] V. van der Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, no. 1, pp. 120–154, 1971.

[7] M. Uppal, Z. Liu, V. Stanković, and Z. Xiong, "Compress-forward coding with BPSK modulation for the half-duplex Gaussian relay channel," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4467–4481, November 2009.

[8] Z. Liu, M. Uppal, V. Stanković, and Z. Xiong, "Compress-forward coding with BPSK modulation for the half-duplex Gaussian relay channel," in *Proc. IEEE*

International Symposium on Information Theory (ISIT), Toronto, Canada, July 2008, pp. 2395 –2399.

[9] M. Uppal, G. Yue, X. Wang, and Z. Xiong, "A rateless coded protocol for half-duplex wireless relay channels," *submitted to IEEE Transactions on Signal Processing*, February 2010.

[10] ——, "A rateless coded protocol for half-duplex wireless relay channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, June 2010.

[11] M. Uppal, Z. Yang, A. Høst-Madsen, and Z. Xiong, "Cooperation in the low power regime for the MAC using rateless multiplexed codes," *IEEE Transactions on Signal Processing, to appear.*

[12] M. Uppal, A. Høst-Madsen, and Z. Xiong, "Practical rateless cooperation in multiple access channels using multiplexed raptor codes," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 671–675.

[13] M. Uppal, Z. Yang, A. Høst-Madsen, and Z. Xiong, "Cooperation in the MAC using frequency division multiplexing," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, South Korea, June 2009, pp. 1373–1377.

[14] M. Uppal, G. Yue, Y. Xin, X. Wang, and Z. Xiong, "A dirty-paper coding scheme for the cognitive radio channel," in *Proc. IEEE International Communications Conference (ICC)*, Captetown, South Africa, May 2010.

[15] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 7, pp. 471–480, July

1973.

[16] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.

[17] T. Cover and J. Thomas, *Elements of Information Theory, 2nd Edition*. Hoboken, NJ: John Wiley & Sons, 2006.

[18] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[19] A. Wyner, "Recent results in the Shannon theory," *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, January 1974.

[20] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 80–94, September 2004.

[21] V. Stanković, A. Liveris, Z. Xiong, and C. Georghiades, "On code design for the general Slepian-Wolf problem and for lossless multiterminal communication networks," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1495–1507, April 2006.

[22] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, vol. 2, San Antonio, TX, December 2001, pp. 1400–1404.

[23] J. Garcia-Frias and W. Zhong, "LDPC codes for compression of multiterminal sources with hidden Markov correlation," *IEEE Communications Letters*, vol. 7,

no. 3, pp. 115–117, March 2003.

[24] D. Schonberg, K. Ramchandran, and S. Pradhan, "Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources," in *Proc. Data Compression Conference (DCC)*, Snowbird, UT, March 2004, pp. 292–301.

[25] A. Wyner, "The rate-distortion function for source coding with side information at the decoder – II: General sources," *Information and Control*, vol. 38, no. 1, pp. 60–80, 1978.

[26] M. Eyuboglu and G. F. Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks – high-rate theory for memoryless sources," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 46–59, January 1993.

[27] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Transactions on Communications*, vol. 38, no. 1, pp. 82–93, January 1990.

[28] Z. Liu, S. Cheng, A. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4358–4379, October 2006.

[29] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *IEEE Transactions on Communications*, vol. 57, no. 2, pp. 376–387, February 2009.

[30] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Probl. Controls and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[31] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[32] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interferences," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3820–3833, November 2005.

[33] Y. Sun, M. Uppal, A. Liveris, S. Cheng, V. Stanković, and Z. Xiong, "Nested turbo codes for the Costa problem," *IEEE Transactions on Communications*, vol. 56, no. 3, pp. 388–399, March 2008.

[34] Y. Sun, Y. Yang, A. Liveris, V. Stanković, and Z. Xiong, "Near-capacity dirty-paper code design: A source-channel coding approach," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3013–3031, July 2009.

[35] U. Erez and S. Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3417–3432, October 2005.

[36] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electronic Letters*, vol. 7, no. 3, pp. 138–139, March 1971.

[37] M. Miyakawa and H. Harashima, "A method of code conversion for a digital communication channel with intersymbol interference," *IEEE Transactions on Communications*, vol. COM-20, no. 8, pp. 774–780, August 1972.

[38] W. Yu, D. Varodayan, and J. Cioffi, "Trellis and convolutional precoding for transmitter-based interference pre-subtraction," *IEEE Transactions on Communications*, vol. 53, no. 7, pp. 1220 – 1230, July 2005.

[39] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups.* New York, NY: Springer-Verlag, 1998.

[40] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.

[41] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4152–4172, December 2005.

[42] H. Chong, M. Motani, and H. Garg, "New coding strategies for the relay channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 2005, pp. 1086–1090.

[43] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, September 2005.

[44] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks, efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062 – 3080, December 2004.

[45] M. Khojastepour, A. Sabharwal, and B. Aazhang, "On capacity of Gaussian 'cheap' relay channel," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, vol. 3, San Francisco, CA, December 2003, pp. 1776 – 1780.

[46] B. Zhao and M. Valenti, "Distributed turbo coded diversity for relay channel," *IEEE Electronics letters*, vol. 39, no. 5, pp. 786–787, May 2003.

[47] M. Janani, A. Hedayat, T. E. Hunter, and A. Nosratinia, "Coded cooperation in wireless communications: Space-time transmission and iterative decoding," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 362–371, February 2004.

[48] Z. Zhang and T. Duman, "Capacity-approaching turbo coding and iterative decoding for relay channels," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1895–1905, November 2005.

[49] ——, "Capacity-approaching turbo coding for half-duplex relaying," *IEEE Transactions on Communications*, vol. 55, no. 10, pp. 1895–1906, October 2007.

[50] A. Chakrabarti, A. De Baynast, A. Sabharwal, and B. Aazhang, "Low density parity check codes for the relay channel," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 280–291, February 2007.

[51] J. Hu and T. Duman, "Low density parity check codes over wireless relay channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3384–3394, September 2007.

[52] P. Razaghi and W. Yu, "Bilayer low density parity check codes for decode and forward in relay channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3723–3739, October 2007.

[53] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 628–643, March 2003.

[54] Z. Liu, V. Stanković, and Z. Xiong, "Practical compress-forward code design for the half-duplex relay channel," in *Proc. IEEE Conference on Information*

Sciences and Systems (CISS), Baltimore, MD, March 2005.

[55] ——, "Wyner-ziv coding for the half-duplex relay channel," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 5, Philadelphia, PA, March 2005, pp. v/1113 – v/1116.

[56] R. Hu and J. Li, "Practical compress-forward in user cooperation: Wyner-Ziv cooperation," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seattle, WA, July 2006, pp. 489–493.

[57] ——, "Exploiting Slepian-Wolf codes in wireless user cooperation," in *Proc. IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, New York, NY, June 2005, pp. 275–279.

[58] A. Chakrabarti, A. De Baynast, A. Sabharwal, and B. Aazhang, "Half-duplex estimate-and-forward relaying: bounds and code design," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seattle, WA, July 2006, pp. 9–14.

[59] B. Djeumou, S. Lasaulce, and A. Klein, "Practical quantize-and-forward schemes for the frequency division relay channel," *EURASIP Journal on Wireless Communications and Networking*, vol. 4, 2007, doi:10.1155/2007/20258.

[60] J. Shea, T. Wong, and B. Choi, *Cooperative Communications for Improved Wireless Network Transmission.* Hershey, PA: IGI Global, 2008, pp. 135–186.

[61] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.

[62] A. Liveris, Z. Xiong, and C. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in *Proc. MMSP-2002*, St. Thomas, US Virgin Islands, December 2002, pp. 53–56.

[63] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd International Symposium on Turbo codes and related topics*, Brest, France, September 2000, pp. 1–8.

[64] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1727–1737, October 2001.

[65] S. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 657–670, February 2001.

[66] S. Chung, G. Forney Jr, T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB ofthe Shannon limit," *IEEE Communications Letters*, vol. 5, no. 2, pp. 58–60, February 2001.

[67] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 564–579, February 1998.

[68] A. Roumy, D. Declercq, and E. Fabre, "Low complexity code design for the 2-user gaussian multiple access channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Chicago, IL, July 2004, p. 481.

[69] A. Amraoui, S. Dusad, and R. Urbanke, "Achieving general points in the 2-user gaussian mac without time-sharing or rate-splitting by means of iterative cod-

ing," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, July 2002, p. 334.

[70] A. Liveris, Z. Xiong, and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, October 2002.

[71] S. Shamai and S. Verdú, "Capacity of channels with uncoded side information," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 587–600, September 2008.

[72] Q. Xu, V. Stankovic, and Z. Xiong, "Layered Wyner-Ziv video coding for transmission over unreliable channels," *Signal Processing*, vol. 86, no. 11, pp. 3212–3225, 2006.

[73] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-checkcodes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[74] A. Ashikhmin, G. Kramer, and S. Ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2657–2673, November 2004.

[75] J. Castura and Y. Mao, "Rateless coding for wireless relay channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1638–1642, May 2007.

[76] X. Liu and T. Lim, "Fountain codes over fading relay channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3278–3287, June 2009.

[77] R. Nikjah and N. Beaulieu, "Achievable rates and fairness in rateless coded relaying schemes," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4439–4444, November 2008.

[78] A. Molisch, N. Mehta, J. Yedidia, and J. Zhang, "Performance of fountain codes in collaborative relay networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, pp. 4108–4119, November 2007.

[79] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

[80] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.

[81] P. Mitran, H. Ochiai, and V. Tarokh, "Space-time diversity enhancements using collaborative communications," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2041 – 2057, June 2005.

[82] M. Luby, "LT codes," in *Proc. IEEE Symposium on the Foundations of Computer Science*, 2002, pp. 271–280.

[83] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Available at http://stanford.edu/ boyd/cvx.

[84] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE,*, vol. 37, no. 1, pp. 10–21, January 1949.

[85] Y. Yao, X. Cai, and G. Giannakis, "On energy efficiency and optimum resource allocation of relay transmissions in the low-power regime," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2917–2927, June 2005.

[86] A. E. Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1545–1561, April 2006.

[87] D. Tse and P. Viswanath, *Fundamentals of wireless communications*. Cambridge,UK: Cambridge University Press, 2005.

[88] A. Avestimehr and D. Tse, "Outage capacity of the fading relay channel in the low-SNR regime," *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1401–1415, April 2007.

[89] D. M. Pozar, *Microwave Engineering*. Hoboken, NJ: John Wiley & Sons, 1998.

[90] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[91] A. Høst-Madsen, M. Uppal, and Z. Xiong, "On outage capacity in the low power regime," *submitted to IEEE Transactions on Information Theory*, May 2010.

[92] B. Wang, J. Zhang, and A. Høst-Madsen, "On the ergodic capacity of MIMO relay channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 29–43, January 2005.

[93] G. Yue, X. Wang, Z. Yang, and A. Høst-Madsen, "Coding schemes for user cooperation in low-power regimes," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2035–2049, May 2008.

[94] Z. Yang, "Cooperation in MAC and relay channels," Ph.D. dissertation, University of Hawaii, 2006.

[95] L. Li, N. Jindal, and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1326–1347, April 2005.

[96] G. Caire, D. Tuninetti, and S. Verdu, "Suboptimality of TDMA in the low-power regime," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 608–620, April 2004.

[97] A. Lozano, A. M. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2527–2544, October 2003.

[98] A. Avestimehr and D. Tse, "Outage-optimal relaying in the low SNR regime," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 2005, pp. 941–945.

[99] L. Xiao, T. Fuja, J. Kliewer, and D. Costello, "Nested codes with multiple interpretations," in *Proc. IEEE Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2006, pp. 851–856.

[100] Z. Yang and A. Høst-Madsen, "Cooperation efficiency in the low power regime (invited paper)," in *Proc. 39th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October/November 2005, pp. 1742–1746.

[101] ——, "Rateless Coded Cooperation for Multiple-Access Channels in the Low Power Regime," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seattle, WA, July 2006, pp. 967–971.

[102] R. Palanki and J. Yedidia, "Rateless codes on noisy channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Chicago, IL, June 2004,

p. 37.

[103] T. Richardson and R. Urbanke, *Modern coding theory*. New York, NY: Cambridge University Press, 2008.

[104] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 3945 –3958, September 2009.

[105] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4707 –4712, October 2008.

[106] ——, "A new achievable rate region for the cognitive radio channel," in *Proc. IEEE International Communications Conference (ICC)*, Beijing, China, May 2008, pp. 1055–1059.

[107] S. Seyedmehdi, J. Jiang, Y. Xin, and X. Wang, "An improved achievable rate region for causal cognitive radio," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Seoul, South Korea, June 2009, pp. 611–615.

[108] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 1361–1391, July 1999.

[109] J. Hou, P. Siegel, L. Milstein, and H. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2141–2155, September 2003.

[110] H. Sankar, N. Sindhushayana, and K. Narayanan, "Design of low-density parity-check (LDPC) codes for high order constellations," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dallas, TX, November/December 2004, pp. 3113–3117.

APPENDIX A

CAPACITY OF A BIMGN CHANNEL

In this appendix, we derive the capacity (3.5) of a BIMGNC. Consider the channel with the output $Y = X + S + Z$, where X is the equiprobable BPSK modulated input with power P, S is the equiprobable BPSK modulated interference with power P_i, and Z is the AWGN with variance σ_2. Let $x = \sqrt{P}$, and $s = \sqrt{P_i}$. The capacity of this channel is given as

$$\begin{aligned} C^{BIMGN} &= I(X;Y) = h(Y) - \frac{1}{2}h\left(Y|X=+x\right) - \frac{1}{2}h\left(Y|X=-x\right) \\ &= -\int_{-\infty}^{\infty} f(y)\log f(y)dy + \frac{1}{2}\sum_{d=\pm x}\int_{-\infty}^{\infty} f(y|X=d)\log f(y|X=d)dy \\ &= \frac{1}{2}\sum_{d=\pm x}\int_{-\infty}^{\infty} f(y|X=d)\log\frac{f(y|X=d)}{f(y)}dy \end{aligned} \tag{A.1}$$

Replacing $f(y) = \frac{1}{2}\sum_{d=\pm x} f(y|X=d)$ in (A.1) we obtain

$$C^{BIMGN} = 1 - \frac{1}{2}\sum_{d=\pm x}\int_{-\infty}^{\infty} f(y|X=d)\log\left(1+\frac{f(y|X=-d)}{f(y|X=d)}\right)dy \tag{A.2}$$

Next, since

$$f(y|X=d) = \frac{1}{2\sqrt{2\pi\sigma^2}}\sum_{m=\pm s}\exp\left(-\frac{(y-d-s)^2}{2\sigma^2}\right), \tag{A.3}$$

we note that $f(y|X=d) = f(-y|X=-d)$. Since the integration in (A.2) is symmetric around zero, we conclude that

$$C^{BIMGN} = 1 - \int_{-\infty}^{\infty} f(y|X=x)\log\left(1+\frac{f(y|X=-x)}{f(y|X=x)}\right)dy. \tag{A.4}$$

Substituting (A.3) in (A.4) gives us the capacity expression (3.5).

APPENDIX B

DERIVATION OF CONDITIONAL PROBABILITY DENSITY OF Y_R GIVEN Y_{D1}

In this appendix, we provide the derivation of the expression (3.14) for the conditional probability density of Y_r given Y_{d1} . Define $x = \sqrt{P_{s1}}$. Then, we have

$$
\begin{aligned}
f\left(y_r|y_{d1}\right) &= \frac{f\left(y_r, y_{d1}\right)}{f\left(y_{d1}\right)} \\
&= \frac{\sum_{d=\pm x} f\left(y_r, y_{d1}|X_{s1}=d\right)}{\sum_{d=\pm x} f\left(y_{d1}|X=d\right)} \\
&= \frac{\sum_{d=\pm x} f\left(y_r, |X_{s1}=d\right) f\left(y_{d1}, |X_{s1}=d\right)}{\sum_{d=\pm x} f\left(y_{d1}|X=d\right)} \qquad \text{(B.1)} \\
&= \frac{\exp\left(-\frac{(y_r-\tilde{c}_{sr})^2}{2} - \frac{(y_d-\tilde{c}_{sd})^2}{2}\right) + \exp\left(-\frac{(y_r+\tilde{c}_{sr})^2}{2} - \frac{(y_d+\tilde{c}_{sd})^2}{2}\right)}{\sqrt{2\pi}\left[\exp\left(-\frac{(y_d-\tilde{c}_{sd})^2}{2}\right) + \exp\left(-\frac{(y_d+\tilde{c}_{sd})^2}{2}\right)\right]} \\
&= \frac{1}{1+\exp\left(-2\tilde{c}_{sd}y_{d1}\right)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y_r-\tilde{c}_{sr})^2}{2}\right) + \\
&\quad \frac{\exp\left(-2\tilde{c}_{sd}y_{d1}\right)}{1+\exp\left(-2\tilde{c}_{sd}y_{d1}\right)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(y_r+\tilde{c}_{sr})^2}{2}\right) \qquad \text{(B.2)}
\end{aligned}
$$

where (B.1) follows from the fact that given X_{s1}, Y_r and Y_{d1} are independent. Substituting the definition of $\zeta\left(y_{d1}\right)$ and $f_g(\cdot)$ in (B.2), we obtain (3.14). ∎

APPENDIX C

PROOF THAT THE DJSCC CHANNELS ARE SYMMETRIC

If X is the binary input to the channel and Y is the output, a channel is symmetric if $f(y|X=0) = f(-y|X=1)$ [73]. We need to prove that both the DJSCC channels, the physical noisy channel, and the virtual correlation channel are symmetric. It is quite clear that the physical noisy channel is symmetric, therefore, we will only consider the correlation channel here. The input to the virtual correlation channel is W and the output is Y_{d1}. We need to prove that $f\left(y_{d1}|W=0\right) = f\left(y_{d1}|W=1\right)$. Using the conditional pdf $f(y_r|y_{d1})$ from (3.14), the function $Q(x) = \int_x^\infty f_g(y)dy$, and $\zeta(y) = \left(1+e^{-2\tilde{c}_{sd}y}\right)^{-1}$, we have

$$\begin{aligned} f\left(y_{d1}|W=0\right) &= \frac{f\left(y_{d1}\right)}{\mathrm{P}(W=0)}\mathrm{P}\left(W=0|y_{d1}\right) \\ &= \frac{f\left(y_{d1}\right)}{\mathrm{P}(W=0)}\left[\zeta(y_{d1})Q(-\tilde{c}_{sr}) + \left(1-\zeta(y_{d1})\right)Q(\tilde{c}_{sr})\right] \\ &= \frac{f\left(y_{d1}\right)}{\mathrm{P}(W=0)}\left[\zeta(y_{d1})Q(-\tilde{c}_{sr}) + \zeta(-y_{d1})Q(\tilde{c}_{sr})\right]. \end{aligned}$$

Also,

$$\begin{aligned} f\left(-y_{d1}|W=1\right) &= \frac{f\left(-y_{d1}\right)}{\mathrm{P}(W=1)}\mathrm{P}\left(W=1|-y_{d1}\right) \\ &= \frac{f\left(-y_{d1}\right)}{\mathrm{P}(W=1)}\left[1-\mathrm{P}(W=0|-y_{d1})\right] \\ &= \frac{f\left(-y_{d1}\right)}{\mathrm{P}(W=1)}\left[1-\zeta(-y_{d1})Q(-\tilde{c}_{sr}) - \zeta(y_{d1})Q(\tilde{c}_{sr})\right] \\ &= \frac{f\left(-y_{d1}\right)}{\mathrm{P}(W=1)}\left[Q(\tilde{c}_{sr}) + Q(-\tilde{c}_{sr}) - \zeta(-y_{d1})Q(-\tilde{c}_{sr}) - \zeta(y_{d1})Q(\tilde{c}_{sr})\right] \\ &= \frac{f\left(-y_{d1}\right)}{\mathrm{P}(W=1)}\left[\zeta(y_{d1})Q(-\tilde{c}_{sr}) + \zeta(-y_{d1})Q(\tilde{c}_{sr})\right]. \end{aligned}$$

Since $f(y_{d1}) = f(-y_{d1})$ (Y_{d1} is symmetric around zero) and $\mathrm{P}(W=0) = \mathrm{P}(W=1) = 0.5$, we have $f\left(y_{d1}|W=0\right) = f\left(y_{d1}|W=1\right)$.

APPENDIX D

PROOF OF PROPOSITION 2

It is straightforward to show that R_{sd} and R_{rd} should lie on the sum-rate side of the MAC capacity region to maximize the rate. Here, we only the address the issue of where exactly should the rates be on the sum-rate side. Let $t \in [0,1]$ be a parameter which indicates the relative positions of R_{sd} and R_{rd} on the sum-rate side, i.e.

$$\begin{aligned} R_{sd} &= tI(X_{s2};Y_{d2}) + \bar{t}I(X_{s2};Y_{d2}|X_r), \quad \text{and} \\ R_{rd} &= \bar{t}I(X_r;Y_{d2}) + tI(X_r;Y_{d2}|X_{s2}). \end{aligned}$$

Notice that the value $t = 0$ corresponds to one corner point where W is decoded first by treating X_{s2} as binary interference. Similarly, $t = 1$ corresponds to the other corner point. All other values of t indicate operation on the non corner points. The proof thus requires us to show that $t = 0$ is the optimum choice.

The overall CF achievable rate (4.12) can then be generalized as

$$R_{CF} = \max_{0\le t\le 1} \left\{ \alpha^*(t) I(X_{s1};W,Y_{d1}) + \overline{\alpha^*(t)} \left[tI(X_{s2};Y_{d2}) + \bar{t}I(X_{s2};Y_{d2}|X_r) \right] \right\}, \quad \text{(D.1)}$$

where the optimum half-duplexing parameter in (4.14), generalized as a function of t is given as

$$\alpha^*(t) = \left(1 + \frac{H(W|Y_{d1})}{\bar{t}I(X_r;Y_{d2}) + tI(X_r;Y_{d2}|X_{s2})} \right)^{-1}. \quad \text{(D.2)}$$

Substituting (D.2) in (D.1), we get

$$R_{CF} = \max_{0\le t\le 1} \left\{ \alpha^*(t) \left[I(X_{s1};Y_{d1}) - H(W|X_{s1}) \right] + \overline{\alpha^*}(t) I(X_r, X_{s2};Y_{d2}) \right\}, \quad \text{(D.3)}$$

Now since $I(X_r;Y_{d2}) \le I(X_r;Y_{d2}|X_{s2})$, $\alpha^*(t)$ in (D.2) is non-decreasing in t. Thus the rate in (D.3) is maximized for either $t = 0$ when $I(X_{s1};Y_{d1}) - H(W|X_{s1}) \le$

$I(X_r, X_{s2}; Y_{d2})$, or $t = 1$ when $I(X_{s1}; Y_{d1}) - H(W|X_{s1}) > I(X_r, X_{s2}; Y_{d2})$. However, notice that when $t = 1$ is the best choice, we have

$$\begin{aligned} R_{CF} &< I(X_{s1}; Y_{d1}) - H(W|X_{s1}) \\ &< C(c_{sd}^2 P) \end{aligned}$$

meaning that it is better not to relay at all. Thus for CF relaying scheme presented in Section 2, Chapter IV, $t = 0$ is the optimum choice, completing the required proof.

APPENDIX E

EXPRESSIONS FOR BUCKET FILLING RATES

Define

$$\mathcal{K}(i,\mathcal{U},\mathcal{S}) \equiv \left\{k \in \{1,\ldots,N\} - \{i\} | \mathcal{D}_k^i \subseteq \mathcal{S}, \mathcal{U} \cap \mathcal{D}_k^i \neq \emptyset\right\},$$

where $\mathcal{D}_j^i = \mathcal{D}_j - \{i\}$. Then for multiplexed coding, the fill rate of a set of users $\mathcal{U} \subseteq \mathcal{S}$ at node i is (with the convention that an empty sum gives zero)

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{|\mathcal{U}|} \log\left(1 + \frac{\sum_{k\in\mathcal{K}(i,\mathcal{U},\mathcal{S})} |c_{ki}|^2 \mathrm{SNR}}{1 + \sum_{k:\mathcal{D}_k^i \not\subseteq \mathcal{S}} |c_{ki}|^2 \mathrm{SNR}}\right) \qquad \text{CDMA},$$

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{N|\mathcal{U}|} \sum_{k\in\mathcal{K}(i,\mathcal{U},\mathcal{S})} \log\left(1 + |c_{ki}|^2 N\mathrm{SNR}\right) \qquad \text{FDMA},$$

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{|\mathcal{U}|} \sum_{k\in\mathcal{K}(i,\mathcal{U},\mathcal{S})} |c_{ki}|^2 \qquad \mathrm{SNR} \to 0.$$

For superposition coding, assuming that a user splits its power equally amongst its decoded set of messages, the fill rate for a set of users $\mathcal{U} \subseteq \mathcal{S}$ at node i is

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{|\mathcal{U}|} \log\left(1 + \frac{\sum_{k=1,k\neq i}^{N} \frac{|\mathcal{U}\cap\mathcal{D}_k^i|}{|\mathcal{D}_k|} |c_{ki}|^2 \mathrm{SNR}}{1 + \sum_{k=1,k\neq i}^{N} \frac{|\mathcal{S}^c\cap\mathcal{D}_k - \mathcal{D}_i|}{|\mathcal{D}_k|} |c_{ki}|^2 \mathrm{SNR}}\right) \qquad \text{CDMA},$$

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{N|\mathcal{U}|} \sum_{k=1,k\neq i}^{N} \log\left(1 + \frac{\frac{|\mathcal{U}\cap\mathcal{D}_k^i|}{|\mathcal{D}_k|} |c_{ki}|^2 N\mathrm{SNR}}{1 + \frac{|\mathcal{S}^c\cap\mathcal{D}_k - \mathcal{D}_i|}{|\mathcal{D}_k|} |c_{ki}|^2 N\mathrm{SNR}}\right) \qquad \text{FDMA},$$

$$r_i(\mathcal{U},\mathcal{S}) = \frac{1}{|\mathcal{U}|} \sum_{k=1,k\neq i}^{N} \frac{|\mathcal{U}\cap\mathcal{D}_k^i|}{|\mathcal{D}_k|} |c_{ki}|^2 \qquad \mathrm{SNR} \to 0.$$

VITA

Name	Momin Ayub Uppal
Education	Doctor of Philosophy (August 2006 – August 2010) Major: Electrical Engineering Texas A&M University, College Station, TX, 77843.
	Master of Science (January 2004 – August 2006) Major : Electrical Engineering Texas A&M University, College Station, TX, 77843.
	Bachelor of Science (August 1998 – May 2002) Major : Electronic Engineering GIK Institute, Swabi, Pakistan.
Research Interests	Coding for cooperative communications, Joint source-channel coding, Dirty-paper coding.
Address	Department of Electrical and Computer Engineering, c/o Dr. Zixiang Xiong Texas A&M University, College Station, TX, 77843.
Email	momin.uppal@gmail.com

The typist for this dissertation was Momin Ayub Uppal.